THE UNIVERSITY *of* EDINBURGH

Investigation of the role of essential proteins in gene silencing at the centromere of *Schizosaccharomyces pombe*

Edward Dobbs

Thesis presented for the Degree of Doctor of Philosophy

University of Edinburgh

2011

Preface

This thesis was composed by me and the research presented is my own unless otherwise stated.

Edward Dobbs

2011

Abstract

The centromeres of eukaryotes have a region on which the kinetochore is assembled, flanked by heterochromatin which provides cohesion between the sister chromatids during cell division. When centromeric heterochromatin is lost chromosomes no longer segregate evenly into the daughter cells during cell division. In the fission yeast *Schizosaccharomyces pombe* (*S. pombe*) RNA interference (RNAi) is responsible for maintaining this heterochromatin. The pathway is part of a feedback loop whereby siRNAs generated from non-coding centromere transcripts are loaded into an Argonaute complex. The siRNAs guide the complex to the homologous centromere repeats in order to recruit Clr4 which modifies histone H3 with the heterochromatin mark H3K9me.

A previous screen to find factors affecting centromere silencing isolated 13 loci termed ***c***entromere: ***s***uppressor of ***p***osition-effect (*csp) 1-13*. Several *csp* mutants have been identified to be RNAi components. In this investigation the *csp6* locus has been identified to be the Hsp70 gene *ssa2*$^{+}$. It has been demonstrated that Argonaute proteins from plants and flies require Hsp70/90 chaperone activity for loading of siRNA. It therefore seems likely that Hsp70 may play a similar role in fission yeast. Genetic and biochemical techniques have been used in this study to investigate if the *csp6* alleles are affecting siRNA loading in *S. pombe*.

RNA Polymerase II (RNAPII) transcribes the pre-siRNA transcripts from the centromere repeats. *csp3* was identified to be an allele of the RNAPII subunit *rpb7*$^{+}$. *rpb7-G150D* was found to cause a silencing defect in the centromeric heterochromatin through a defect in transcription. Another RNAPII mutation, *rpb2-m203*, was found to have strong silencing defects caused by an unidentified non-transcriptional role in RNAi-mediated heterochromatin formation at the centromere. In order to gain more insight into the role of RNAPII in heterochromatin assembly I performed a screen in which the subunits *rpb3* and *rpb11* were subjected to random mutagenesis. Several mutants were isolated and characterisation of phenotypes regarding heterochromatin at the centromere has

been carried out for nine of the mutants. As a result a novel phenomenon of RNAi-independent silencing at the centromere has been discovered.

Abbreviations

ADP	Adenosine Di-Phosphate
AIR	Amino-Imidazole Ribonucleotide
ARC	Argonaute siRNA Chaperone
ars	Autonomous Replication Sequence
ATP	Adenosine Tri-Phosphate
bp	Base Pairs
CATD	CENP-A Targeting Domain
cDNA	Complementary DNA
CENP-A	Centromere Protein A
ChIP	Chromatin Immunoprecipitation
cloNAT	Nourseothricin
ClrC	Clr4 Complex
cnt	Central Core Region
Co-IP	Co-Immunoprecipitation
csp	Centromere Suppressor Of Position Effect
CTD	C-Terminal Domain
DNA	Deoxyribonucleic Acid
dNTP	Deoxyribonucleotide Triphosphate
dsRNA	Double Stranded RNA
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic Acid
EMS	Ethyl Methanesulfanate
ER	Endoplasmic Reticulum
FOA	Fluoroorotic Acid
G418	Geneticin
GAII	Genome Analyser II
GAIIx	Genome Analyser IIx
GBD	Gtpase Protein Binding Domain
GFP	Green Fluorescent Protein
GTF	General Transcription Factor
H3K4me	Histone H3 Lysine4 Methylation
H3K9ac	Histone H3 Lysine9 Acetylation
H3K9me	Histone H3 Lysine9 Methylation
HA	Haemagglutinin Tag
HAT	Histone Acetyltransferases
HDAC	Histone Deacetylases
HEPES	4-(2-Hydroxyethyl)-1-Piperazineethanesulfonic Acid
HFD	Histone Fold Domain
HIRA	Histone Regulatory Complex A

HJURP	Holliday Junction Recognition Protein
HMT	Histone Methyltransferases
Hop	Hsp70/Hsp90 Organising Protein
HP1	Heterochromatin Protein 1
HRP	Horseradish Peroxidase
Hsp	Heat Shock Protein
HULC	Histone H2B Ubiquitin Ligase Complex
imr	Inner or Imperfect Repeat Region
IRC	Centromeric Inverted Repeats
JmjC	Jumonji Domain, C-Terminal Part
kb	Kilo-Bases
LB	Lysogeny Broth
Maq	Mapping And Assembly With Quality
Mb	Mega-Bases
ME	Mealt Extract
Minus URA	Minus Uracil
miRNA	micro RNA
mRNA	messenger RNA
MT	Microtubules
NBD	N-Terminal Nucleotide Binding Domain
ncRNA	Non-coding RNA
NEF	Nucleotide Exchange Factor
ORF	Open Reading Frame
otr	Outer Repeat Region
PBS	Sodium Perborate
PcG	Polycomb Group
PcG	Polycomb Group
PCR	Polymerase Chain Reaction
PEG	Polyethylene Glycol
PEV	Position-Effect Variegation
PIC	Pre-Initiation Complex
piRNA	Piwi-interacting RNA
PMG	Pombe Minimal Glutamate Media
PMSF	Phenylmethanesulfonylfluoride
PNK	Polynucleotide Kinase
priRNA	Primal RNA
PTGS	Post-Transcriptional Gene Silencing
qPCR	Quantitative Polymerase Chain Reaction
qRT-PCR	Quantitative Reverse Transcription PCR
RdRC	RNA-Dependent RNA Polymerase Complex
RdRP	RNA-Dependent RNA Polymerase
RISC	RNA-Induced Silencing Complex

RITS	RNA-Induced Initiation Of Transcription
RNAi	RNA interference
RNAP	RNA Polymerase
RNAPII	RNA Polymerase II
RNAPIII	RNA Polymerase III
RNAPIV	RNA Polymerase IV
rRNA	Ribosomal RNA
RT-PCR	Reverse Transcriptase Polymerase Chain Reaction
SBD	C-Terminal Substrate Binding Domain
SDS	Sodium Dodecyl Sulfate
SDS-PAGE	SDS polyacrylamide gels
SHREC	Snf2/HDAC-Containing Repressor Complex
siRNA	short interfering RNA
SNP	Single Nucleotide Polymorphism
SOC	Super Optimal broth with Catabolite repression
SR	Steroid Hormone Receptor
SSC	Saline-Sodium Citrate
ssRNA	Single Stranded RNA
TBE	Tris/Borate/EDTA
TBZ	Thiabendazole
TE	Transposable Element Or Tris-Edta
TES	N-Tris(Hydroxymethyl)Methyl-2-Aminoethanesulfonic Acid
TF	Transcription Factor
TFIIIC	Transcription Factor IIIC
TFIIS	Transcription Factor IIS
TGS	Transcriptional Gene Silencing
TRAMP	Trf4/Air2/Mtr4P Polyadenylation Complex
tRNA	Transfer RNA
ts	Temperature Sensitive
TUK	Transcripts Under The Kinetochore
Y2H	Yeast Two-Hybrid
YES	Yeast Extract-Sucrose
Zn BD	Zinc Binding Domain

Table of Contents

List of Figures and Tables

Chapter 1 – Introduction

1.1. Heterochromatin

1.1.1 The nature of chromatin

The uncoiled DNA content of a human diploid cell is measured in metres, but the cell itself is measured in micrometres. This simple fact dictates the need for massive compaction of DNA within the cell, and the evolutionary solution to this problem is to use proteins to fold the DNA to form a structure called chromatin. Histone proteins form an octameric complex around which DNA is wound to produce the "beads on a string" structure seen by electron microscopy when chromatin is decondensed (Oudet et al., 1975). One unit of this histone protein/DNA complex is known as a nucleosome, and it consists of the four histone proteins: H2A, H2B, H3 and H4, in duplicate, associated with approximately 146 bp of DNA (Kornberg, 1974). The histones are highly basic, globular proteins which have N-terminal tails that extend outside of the nucleosome structure. These histone tails, particularly those of H3 and H4, can be modified through methylation, acetylation, phosphorylation or ubiquitination of specific amino acids, providing an epigenetic source of gene regulation popularly known as the "Histone Code" (Jenuwein and Allis, 2001). An example of this code can be seen in Figure 1.1. These histone modifications involve a host of enzymes such as: Histone Methyltransferases (HMTs), Histone Demethylases, Histone Acetyltransferases (HATs), Histone Deacetylases (HDACs).

Traditionally, chromatin has been categorised into two types; heterochromatin and euchromatin. This initial characterisation was based on cytological studies which described the distinctive light (relatively diffuse euchromatin) and dark (highly compacted heterochromatin) banding seen on chromosomes under a microscope upon staining (Brown, 1966). Heterochromatin and euchromatin have more recently been redefined based on low and high gene expression, respectively. Heterochromatin has been found to be gene poor with very little transcriptional activity whereas euchromatin

contains the bulk of the genes and is generally transcriptionally active. As we learn more about chromatin, however, this simple black and white definition has been blurred. Heterochromatin is not in fact as transcriptionally inert as previously thought – studies in fission yeast have shown that transcription through heterochromatin is required as part of its structural regulation (Djupedal and Ekwall, 2008). The chromatin domains are also defined at the molecular level via specific histone modifications and associated chromatin proteins. Heterochromatin can be defined by the presence of Histone H3 Lysine9 methylation (H3K9me) and Heterochromatin Protein 1 ($HP1^{Swi6}$), and, in metazoa and plants, also by DNA methylation. Likewise euchromatin has its own characteristic hallmarks, namely Histone H3 Lysine9 acetylation (H3K9ac) and Histone H3 Lysine4 methylation (H3K4me) (Fischle et al., 2003).

Heterochromatin itself can now be defined into two different categories (Arney and Fisher, 2004); constitutive heterochromatin which are regions of a chromosome that are always heterochromatic (eg telomeres and centromeres), and facultative heterochromatin which is a region of heterochromatin that can be switched between expressed or silent states (eg due to dosage compensation).

There is a low level of transcription at heterochromatic loci, these regions are very gene poor, and expression of inserted marker genes is generally suppressed. Thus heterochromatin has often been dubbed 'silent chromatin'. It has been known that heterochromatin has the ability to spread into surrounding chromatin regions since the discovery that a euchromatic gene inserted into a heterochromatic region becomes silenced, a phenomenon known as position-effect variegation (PEV) (Eissenberg, 1989). Much of the work to identify factors controlling PEV was done in *Drosophila melanogaster* however, since the discovery of PEV in *S. pombe* (Allshire et al., 1994), *S. pombe* has also become an important organism for studying heterochromatin formation. Spreading of heterochromatin requires the recruitment of histone deacetylases (HDACs) as well as histone methyltransferases (HMTs) to provide the H3K9me marks that the heterochromatin proteins bind to via their chromodomains (Yamada et al., 2005), (Zhang et al., 2008), (Kagansky et al., 2009).

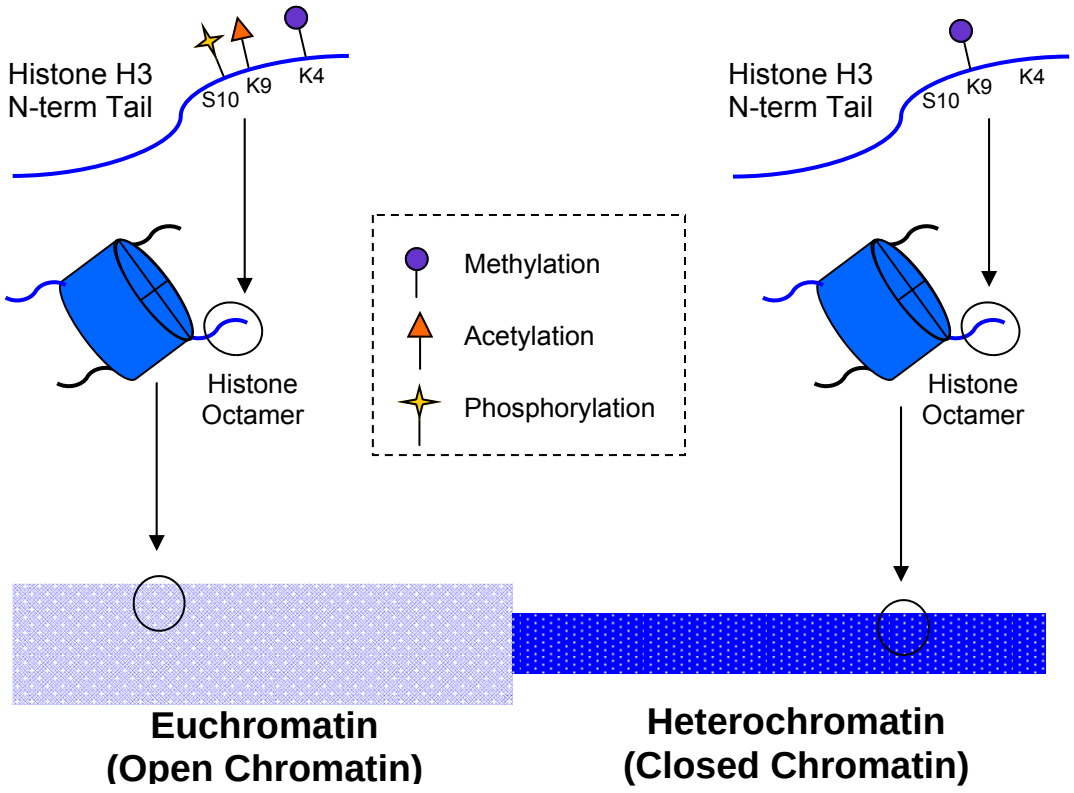


Figure 1.1 – An example of the histone code. A selection of the most well understood modifications of histone protein H3. These modifications distinguish between different chromatin states of the chromosome, the histone H3 tails having different region-specific modifications in euchromatin and in heterochromatin.

In order to prevent spreading of heterochromatin which would cause genes that are normally expressed to be silenced, cells have developed several mechanisms of regulation. tRNA genes and their flanking sequences have been shown to provide a barrier to heterochromatin spreading in yeast, most likely because of their recruitment of RNAPIII and its cofactors (Scott et al., 2006), (Donze and Kamakaka, 2001). The RNAPIII cofactor TFIIIC, in the absence of RNAPIII, has been shown to be recruited to inverted repeats (IR-l and IR-R) at the mating type locus of *S. pombe* to provide barriers to heterochromatin spreading (Thon et al., 2002), (Noma et al., 2006). Deletion of one of the centromeric inverted repeats (IRCs) located at the boundary between heterochromatin and euchromatin in *S. pombe* (Cam et al., 2005) results in loss of barrier function suggesting that there is also a TFIIIC/RNAPIII-independent

barrier mechanism (Noma et al., 2006). The JmjC protein Epe1 is required for maintaining the correct heterochromatin boundaries in *S. pombe*, with heterochromatin domains expanding or retracting in an *epe1*Δ mutant (Trewick et al., 2007).

1.1.2 Functions of heterochromatin

Over the last twenty years advances in our understanding of heterochromatin function have changed our view of this silent chromatin from a genetic junkyard into a subtle regulator of cellular processes. We now know that heterochromatin has both regulatory roles in gene expression, and structural roles in genome organization.

One regulatory role of heterochromatin is in stably silencing particular loci which are important during cellular differentiation in multi-cellular organisms or important for correct protein dosage in imprinting and X-chromosome inactivation. This facultative heterochromatin is formed by recruitment of the Polycomb group (PcG) proteins via diverse nucleation signals and can be stably propagated through mitosis or meiosis (Eun et al., 2010), (Kalantry, 2011).

Inactivation of the X chromosome is a dosage compensation mechanism in female mammals which have double the number of X chromosome genes due to their XX genotype compared to the male XY. The paternal X chromosome is inactivated during early embryonic development by expression of the Xist RNA gene; resulting in coating of the inactivating X chromosome and consequent recruitment of the PcG proteins which modify the chromatin with heterochromatin marks (Kalantry, 2011). In later development of the embryo the choice of which X chromosome is inactivated becomes random, resulting in mosaicism of the organism's cells.

One of the most surprising results of the first completed genome sequencing projects is that as much as 40% of the genome of complex eukaryotes is made up of transposable elements (TEs) (Lander et al., 2001), (Waterston et

al., 2002). TEs, for example retroviral sequences, integrate their genes into the host's genome where they can propagate and cause further random integration of their sequences elsewhere in the genome. These "jumping genes" cause mutation of the host genome and chromosomal rearrangements (Curcio and Derbyshire, 2003) although they are also recognized as important sources of evolutionary variation (Kidwell and Lisch, 1997). Packaging into heterochromatin silences TE expression and thereby suppresses their mutagenic activity, helping to maintain genome integrity.

Heterochromatin is also known to act as an inhibitor of recombination. An example of this role in genome regulation can be seen at the mating-type region of *S. pombe*. The mating-type locus contains three genes which are responsible for mating-type switching, the transcriptionally active *mat1* gene and the silenced *mat2-P* and *mat3-M* genes which are stored in heterochromatin and act as donor sequences for the *mat1* locus (Egel, 2003). The mating-type is determined by which sequence is expressed at the *mat1* locus: either an active copy of the *mat2-P* sequence to produce the Plus mating-type or the *mat3-M* sequence to produce the Minus mating-type. It is therefore important for *S. pombe* meiotic control that the *mat2-P* and *mat3-M* regions are transcriptionally repressed and that the sites do not undergo recombination into an active region, aside from the *mat1* region. Normally the *mat2-P*/*mat3-M* loci exhibit at least 1000 fold less recombination than expected of a normal locus (Egel, 1984) and when heterochromatin is lost, inhibition of recombination is also lost (Thon and Klar, 1992). As discussed below, heterochromatin associated with pericentromeric repeat sequences both suppresses recombination between repeats (Ellermeier et al., 2010) and contributes to the structural organization of the functional centromere.

1.2 Centromeres

1.2.1 Centromere function

The centromere can be seen cytologically as the region of the chromosome at which there is a visible constriction and the position at which sister chromatids are joined. It is at the centromere that the kinetochore, which is the proteinaceous link between the spindle microtubules and the chromosome, is assembled. A functional centromere is essential for equal segregation of DNA into daughter cells, and thus defects in centromere function can cause missegregation of the chromosomes and aneuploid daughter cells. Aneuploidy, the presence of the incorrect number of chromosomal copies in a cell, is the cause of numerous diseases such as Down Syndrome (Lejeune et al., 1959) and Prader Willi Syndrome (Cassidy et al., 1992) and may be important in the progression of cancer (Pellman, 2007).

Considering the vital role that the centromere plays in ensuring stable propagation of the genome, it is surprising that the DNA sequence of centromeres is not at all conserved across species, and in fact the centromere sequence varies from chromosome to chromosome within a species (Willard, 1985), (Wood et al., 2002), (Sanyal et al., 2004). Further evidence of the apparent unimportance of centromere DNA sequence is that neocentromeres can form on non-centromeric sequences (Warburton, 2004). Taken together these observations strongly suggest that centromeres are propagated via an epigenetic mechanism, most likely through the centromere's unique chromatin proteins.

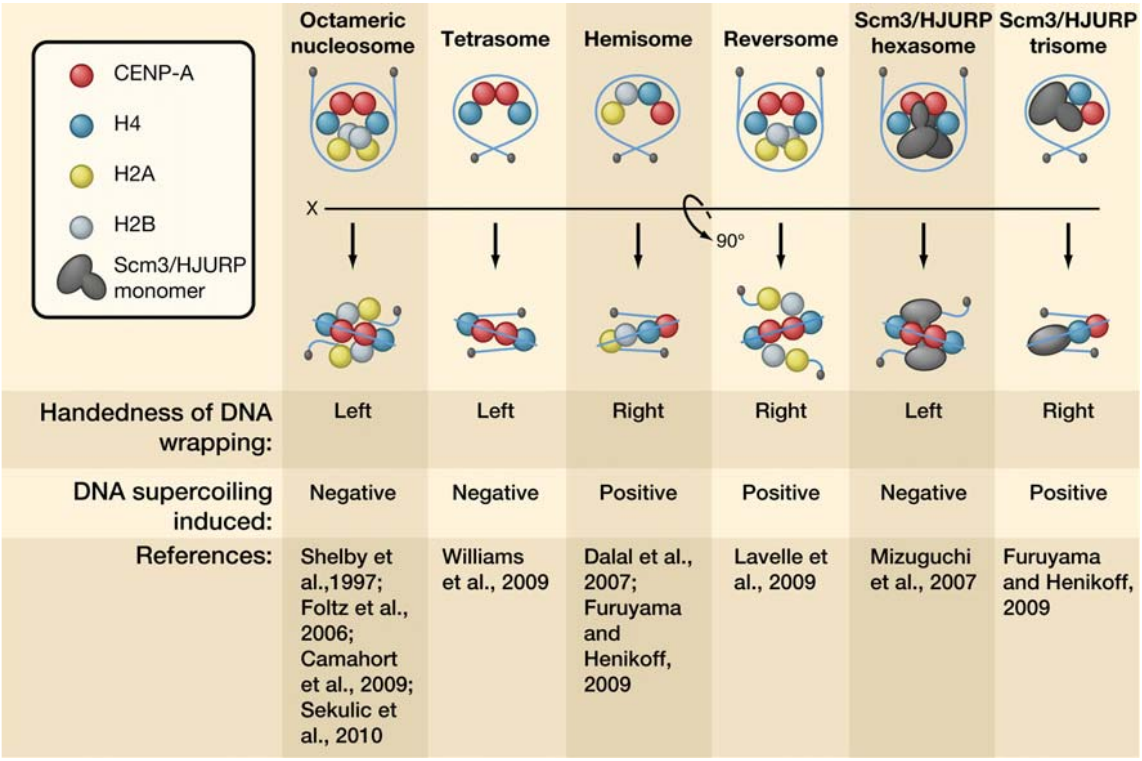
The chromatin structure of the centromere is quite unique in comparison to heterochromatin and euchromatin. Micrococcal nuclease digestion of the centromeric central core in fission yeast produces a smear pattern rather than the typical ladder pattern seen at other chromatin regions (Polizzi and Clarke, 1991). In metazoans the region is made up of blocks of CENP-A nucleosomes amidst blocks of H3-containing nucleosomes in this region (Blower et al., 2002). The central core of fission yeast centromeres is also made up of both CENP-A and

canonical H3, although the exact distributions have not been proven (Allshire and Karpen, 2008). The histone modification pattern of the H3 histones in centromeric chromatin is also unique, with H3K4me2 present (a known modification at euchromatin) but in the absence of any other hallmark modifications of either euchromatin or heterochromatin (Sullivan and Karpen, 2004).

Although the centromere architecture of eukaryotes also varies quite widely, common features are seen: repetitive DNA sequences and an AT-rich central core area packaged in chromatin containing the histone H3 variant, CENP-A^{CenH3,Cnp1}, which is surrounded by heterochromatin. CENP-A, a histone H3 variant found only at the centromere, is vital for assembly of the kinetochore (Collins et al., 2005), and it is thought that CENP-A provides the epigenetic mark for propagation of the centromere. Like other histone proteins, CENP-A contains a histone fold domain (HFD), although uniquely, its HFD contains a CENP-A Targeting Domain (CATD) which is responsible for localizing CENP-A to the centromere (Vermaak et al., 2002), (Black et al., 2004). CENP-A proteins are not well conserved across species when compared to the highly conserved canonical histone proteins and in particular the N-terminal tail of the CENP-A protein has a very diverse sequence across species (Henikoff and Dalal, 2005).

Currently there is great controversy over the exact components that make up a CENP-A nucleosome with new experiments providing seemingly contradictory results (Black and Cleveland, 2011). The theories proposed range from a classical octameric structure, similar to the canonical H3-containing nucleosomes (Camahort et al., 2009) to a tetrameric structure containing only single copies of CENP-A, H4, H2A and H2B (Dalal et al., 2007) and, in *S. cerevisiae*, a hexameric structure in which H2A and H2B have been replaced by two copies of the non-histone protein SCM3 (Mizuguchi et al., 2007). The main details of the structural models are summarised in Figure 1.2. Recently the first crystal structure of a human CENP-A nucleosome was published (Tachiwana et al., 2011) showing a classic octameric structure where H2A, H2B, H4 and CENP-A are present in duplicate.

In *S. pombe* the Scm3 protein, the orthologue of which is part of the CENP-A nucleosome in *S. cerevisiae*, has instead been implicated in deposition of CENP-A (Williams et al., 2009), (Pidoux et al., 2009). The highly conserved proteins Mis16$^{RbAp46/RbAp48}$ and Mis18$^{hMis18\alpha/\beta}$ form a complex which is required for the correct localisation of both Scm3 and consequently CENP-A to the centromere (Williams et al., 2009), (Pidoux et al., 2009). In agreement with this chaperone role for Scm3 the human orthologue HJURP has also recently been shown to deposit CENP-A at the centromere (Dunleavy et al., 2009), (Foltz et al., 2009).



	Octameric nucleosome	Tetrasome	Hemisome	Reversome	Scm3/HJURP hexasome	Scm3/HJURP trisome
Handedness of DNA wrapping:	Left	Left	Right	Right	Left	Right
DNA supercoiling induced:	Negative	Negative	Positive	Positive	Negative	Positive
References:	Shelby et al.,1997; Foltz et al., 2006; Camahort et al., 2009; Sekulic et al., 2010	Williams et al., 2009	Dalal et al., 2007; Furuyama and Henikoff, 2009	Lavelle et al., 2009	Mizuguchi et al., 2007	Furuyama and Henikoff, 2009

Figure 1.2 – The current models for CENP-A nucleosome structure. This figure has been taken directly from (Black and Cleveland, 2011) in which it appears as Figure 2

The kinetochore is the protein complex that mediates the attachment of microtubules to chromosomes during mitosis. The kinetochore complex is made up of >80 proteins (Cheeseman and Desai, 2008) which can be seen by EM as

two defined regions known as the inner and outer kinetochore (McEwen et al., 1998). The inner kinetochore is assembled on CENP-A providing a platform for the outer kinetochore assembly through which the microtubule attachments are made. The kinetochore has been studied a great deal in *S. cerevisiae*, *S. pombe* and *H. sapiens* and the majority of the proteins that make up the kinetochore are conserved from yeast to humans (Musacchio and Salmon, 2007), (Cheeseman and Desai, 2008), (Welburn and Cheeseman, 2008).

1.2.2 Centromeres in the fission yeast *S. pombe*

The fission yeast *S. pombe* is a useful model for investigation of centromeres because the centromeres have similar features to the more complex eukaryotes whilst providing a genetically tractable system. *S. pombe* has three centromeres which share the same basic structure: heterochromatic outer repeats flanking the CENP-A chromatin region which is made up of inner repeats either side of the central core (Figure 1.3) (Partridge et al., 2000). The outer repeats contain two elements, dg and dh, which vary in their number at each centromere. The inner repeats contain tRNA genes which act as boundaries between heterochromatin and CENP-A chromatin (Scott et al., 2006) while the central core is made up solely of CENP-A chromatin. The kinetochore complex is assembled on the CENP-A chromatin within the central core and inner repeats of the *S. pombe* centromeres.

Research on the centromeres of fission yeast over the last decade has resulted in a rapid development of our knowledge of the role of the heterochromatin domains flanking the CENP-A chromatin. Disruption of pericentromeric heterochromatin causes lagging chromosomes at anaphase due to missegregation of sister chromatids. The Heterochromatin Protein (HP)1 homolog Swi6 acts as a platform for the recruitment of the Cohesin complex which holds sister chromatids together during anaphase until correct biorientation is achieved (Bernard, 2001). Missegregation of the sister chromatids occurs when Swi6 is delocalised from the centromeric outer repeats,

disrupting Cohesin localisation and thus losing cohesion between sister chromatids causing aberrant microtubule attachments to occur (Figure 1.4).

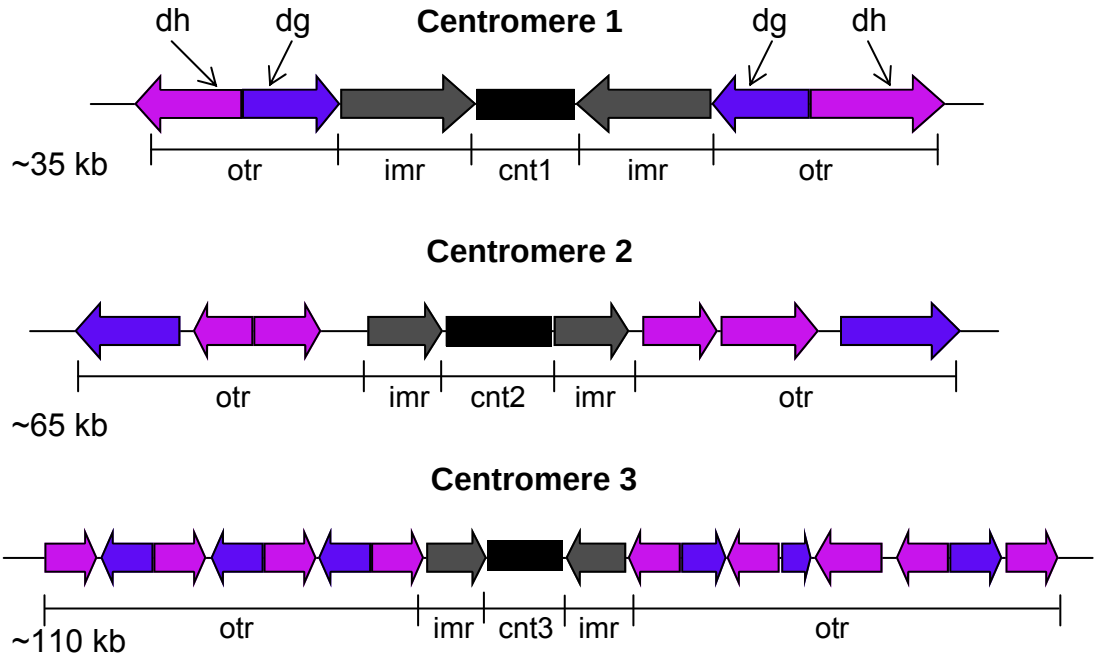


Figure 1.3 – The centromeres of *S. pombe*. *S.pombe* has three centromeres which vary in size and are inversely proportional to the size of the chromosomes. The outer (otr) repeats are packaged in heterochromatin and they are made up of two repetitive elements, dg and dh. The inner or imperfect (imr) repeats contain boundaries between heterochromatin and CENP-A chromatin regions. The central core (cnt) region is made up of CENP-A chromatin which denotes the position of kinetochore assembly.

Pericentromeric heterochromatin is also important for the localisation of CENP-A to the centromere in *S. pombe*. The minimal sequence for generation of a stable centromere on minichromosomes in *S. pombe* includes part of the pericentromeric repeats along with the central core (Takahashi et al., 1992). More recent experiments have shown that heterochromatin in close proximity to the central core is required specifically for the establishment of CENP-A onto naked centromere sequence but not for maintenance of the CENP-A chromatin (Folco et al., 2008). The replacement of the outer repeats on minichromosomes with synthetic heterochromatin also shows that it is the proximity of

heterochromatin, rather than a particular DNA sequence or RNAi factors, close to the central core which allows establishment of CENP-A (Kagansky et al., 2009).

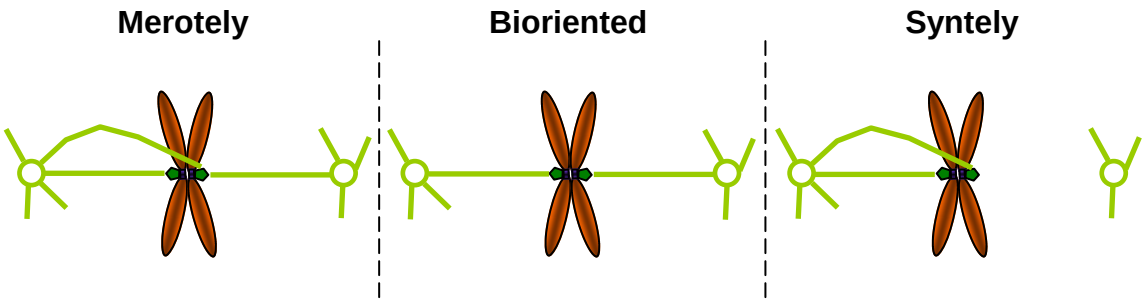


Figure 1.4 – The three main types of spindle attachment during mitosis. Microtubules (MTs) make attachments to the kinetochore located at the centromere of the chromosomes. In order for equal segregation of DNA the sister chromatids must be bioriented (attached to MTs from opposing sides of the spindle). Segregation defects occur when the chromatids have merolic or syntelic MT attachments resulting in unequal inheritance of DNA to daughter cells, lagging chromosomes and/or DNA breaks.

More recently pericentromeric heterochromatin has also been found to inhibit meiotic recombination in the region (Ellermeier et al., 2010). This is important because a defect in meiotic recombination can result in chromosome loss. It was found that the RNAi pathway and the H3K9 methyltransferase Clr4 are particularly important for repression of meiotic recombination.

There are three distinct constitutive heterochromatic regions on the *S. pombe* genome; the telomeres, the pericentromere and the mating-type (mat) locus (Figure 1.5).

RNA interference (RNAi)-directed chromatin modification has been shown to play an important role in the establishment of heterochromatin at all three loci (Hall et al., 2002). Only maintenance of the pericentromeric heterochromatin is dependent on the RNAi pathway however because the mat locus and telomeres have redundant pathways, which run parallel to the RNAi pathway. At the mat locus the Atf1/Pcr1 proteins provide a redundant mechanism for heterochromatin maintenance (Jia et al., 2004) by recruiting the HDACs Clr6 and Clr3 which stimulate heterochromatin spreading across the

locus (Kim et al., 2004), (Yamada et al., 2005). Taz1 nucleates heterochromatin redundantly at the telomeres (Kanoh et al., 2005) by binding to the subtelomeric region and recruiting Clr4 via an unknown mechanism.

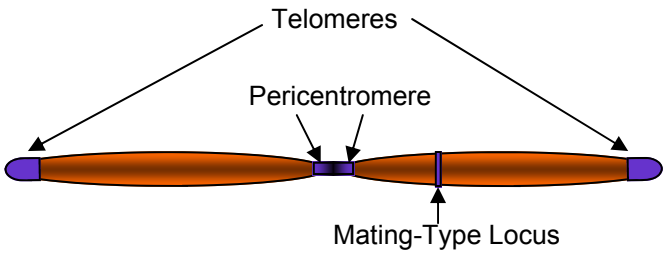


Figure 1.5 – The three main regions of heterochromatin on a *S. pombe* chromosome

1.3 RNA interference (RNAi)

1.3.1 RNAi performs multiple roles in the cell

RNA interference (RNAi) causes silencing of genes at either the transcriptional level (TGS), though recruitment of heterochromatin, or at the post-transcriptional level (PTGS) through prevention of mRNA translation or degradation of mRNA. RNAi effector complexes are guided to their RNA targets by small regulatory RNAs which are homologous to the target RNA. During the 1990s this gene silencing pathway was discovered independently in fungi as Quelling (Romano and Macino, 1992), in plants as Post-Transcriptional Gene Silencing (PTGS) (Dougherty and Parks, 1995) and in nematodes (Fire et al., 1998), flies (Kennerdell and Carthew, 1998) and mammals (Elbashir et al., 2001) as RNAi. Since then the RNAi pathway has been shown to have at least three distinct roles in eukaryotes (Almeida and Allshire, 2005); 1) Defence against transposable elements and viruses, 2) Post-Transcriptional Gene Silencing (PTGS) and 3) Transcriptional Gene Silencing (TGS) and Heterochromatin Formation.

Dicer is a core RNAi protein which is required for generating the small regulatory RNAs (short interfering [si]RNA and micro [mi]RNA) from longer RNA chains. Dicer contains two RNase III domains which provide the exonuclease activity to cut long precursor RNA into small regulatory RNA (MacRae et al., 2006). The Dicer proteins also have a PAZ domain which can recognise and bind the characteristic 3' overhangs on the end of small RNA produced by Dicer (Lingel et al., 2004). The activity of the RNaseIII and PAZ domains and their relative distance apart within Dicer's tertiary structure have led to the theory that Dicer acts as a molecular ruler which ensures the consistent length of the small RNAs produced by Dicer (MacRae et al., 2006).

All RNAi pathways are characterised by the binding of small regulatory RNAs by members of the Argonaute (Ago) protein family. The Argonaute proteins are made up of four domains that are required in all eukaryotic RNAi pathways; the N-terminal, the PAZ domain, the Mid domain and the PIWI domain. The Argonaute PAZ domain is involved in recognition and binding of small RNA (Ma et al., 2004), (Lingel et al., 2004) and it is also found in another core RNAi protein, the RNaseIII enzyme Dicer. The PIWI domain contains an RNase-H-like catalytic group which provides the Argonautes' slicer activity (Song et al., 2004), (Irvine et al., 2006). This slicing activity is required for efficient maturation of double-stranded siRNA into the single-stranded mature siRNA, although slicing is not required for miRNA maturation (Matranga et al., 2005). Slicing is also required for PTGS where the Argonaute protein in RISC causes degradation of the mRNA homologous to the Argonaute-bound siRNA (Irvine et al., 2006). In addition to these roles piwi-interacting (pi)RNA levels in flies are also amplified through the slicing activity of Argonaute family proteins (Brennecke et al., 2007).

1.3.2 RNAi in Post-Transcriptional Gene Silencing (PTGS)

There are two types of PTGS which have differing modes of action depending on the source/nature of the small regulatory RNA: (1) translational repression by micro (mi)RNAs which can have several mismatched nucleotides compared

to the target transcript (2) mRNA degradation by short interfering (si)RNAs or miRNAs which have identical sequence to their target. The exact mechanism of miRNA-mediated translational repression is still poorly understood, with some evidence for repression at the level of either translation initiation or post-initiation stages (Fabian et al., 2010). The mechanism of mRNA degradation however appears to be more straightforward, through cleavage of mRNA by the Argonaute's slicing activity (Liu and Hendrickson, 2007).

miRNA and siRNA are defined by their source RNA structure, where miRNAs are produced by specific cleavage of ssRNAs folded into a hairpin structure, whereas siRNAs are produced from linear dsRNA (Figure 1.6). Both of these small RNA species are cut down to mature small RNA of 21-28 nucleotides in length by the RNAse III activity of Dicer (Meister and Tuschl, 2004). Upstream of Dicer, the nuclease Drosha also carries out some initial processing of the pre-miRNA transcript. In both cases the mature small RNA is loaded into an Argonaute protein which binds other proteins to form the effector complex; RNA-induced Silencing Complex (RISC) for siRNA (Hammond 2000) and miRNA-containing Ribonucleoprotein Particle (miRNP) for miRNA (Mourelatos et al., 2002).

In addition to Dicer and a RISC-like effector, some RNAi pathways require an RNA-dependent RNA Polymerase (RdRP). An RdRP plays an essential role in the systemic spread of antiviral silencing in plants and nematodes, through amplification of the siRNA levels (Baulcombe, 2004). The genomes of *Drosophila* and mammals do not appear to encode RdRPs, implying that such amplification is not required for all RNAi pathways. It is commonly proposed that the RNAi pathway originally evolved as a viral defense mechanism to protect ancient eukaryotes against viral nucleic acids, as is one of its roles in modern eukaryotes.

As well as siRNAs and miRNAs there is a third distinct class of small regulatory RNAs known as Piwi-interacting (pi)RNA. These 24-32nt small RNAs are generated in a Dicer-independent manner which is currently not well understood (Siomi et al., 2011). The piRNAs are loaded into the Argonaute subfamily of Piwi proteins which degrade homologous RNAs by slicing (Saito et

al., 2006), (Vagin et al., 2006), the products of which are used to amplify piRNA levels (Brennecke et al., 2007), (Gunawardane et al., 2007). piRNA, originally known as repeat-associated (rasi)RNA, is used to silence transposable elements (TEs) and repetitive DNA in the germlines of animals. The PTGS mode of silencing by piRNAs is now well accepted, however there is evidence that there is also a transcriptional gene silencing (TGS) role for piRNAs in TE suppression in the male germline cells of mice (Aravin et al., 2007), (Kuramochi-Miyagawa et al., 2008).

Probably the best known effect of RNAi is the PTGS process of mRNA degradation which is controlled by the Argonaute-containing effector complex RISC. PTGS is now widely applied in research as a means to dramatically knockdown a specific target transcript using homologous dsRNA which is processed into short interfering (si)RNA, loaded into RISC and used as a template to target the homologous transcript. This method is now an invaluable tool in genetic manipulation in higher eukaryotes where alteration of the genome is difficult.

While The RISC complex carries out PTGS in the cytoplasm, the *R*NA-induced *I*nitiation of *T*ranscription (RITS) Complex, which is a RISC variant, is based in the nucleus where it carries out transcriptional gene silencing (TGS). TGS is a mechanism of gene regulation whereby the gene is assembled into heterochromatin via recruitment of histone methyltransferases to the locus by the RNAi pathway. Although TGS has been a well documented phenomenon it was not until 2001 that it was found to be under the control of RNAi (Jones et al., 2001).

The heterochromatin at the centromeres of *S. pombe* was shown to be formed by TGS by Volpe et al in 2002 (Volpe et al., 2002) and since then *S. pombe* has been widely used to develop our knowledge of the pathway. *S. pombe* is a useful model organism to test TGS because it has only a single copy of each of the main RNAi proteins, whereas in other organisms there are several copies of the Argonaute and Dicer proteins. TGS will be discussed in greater detail, using the *S. pombe* model, in the next section.

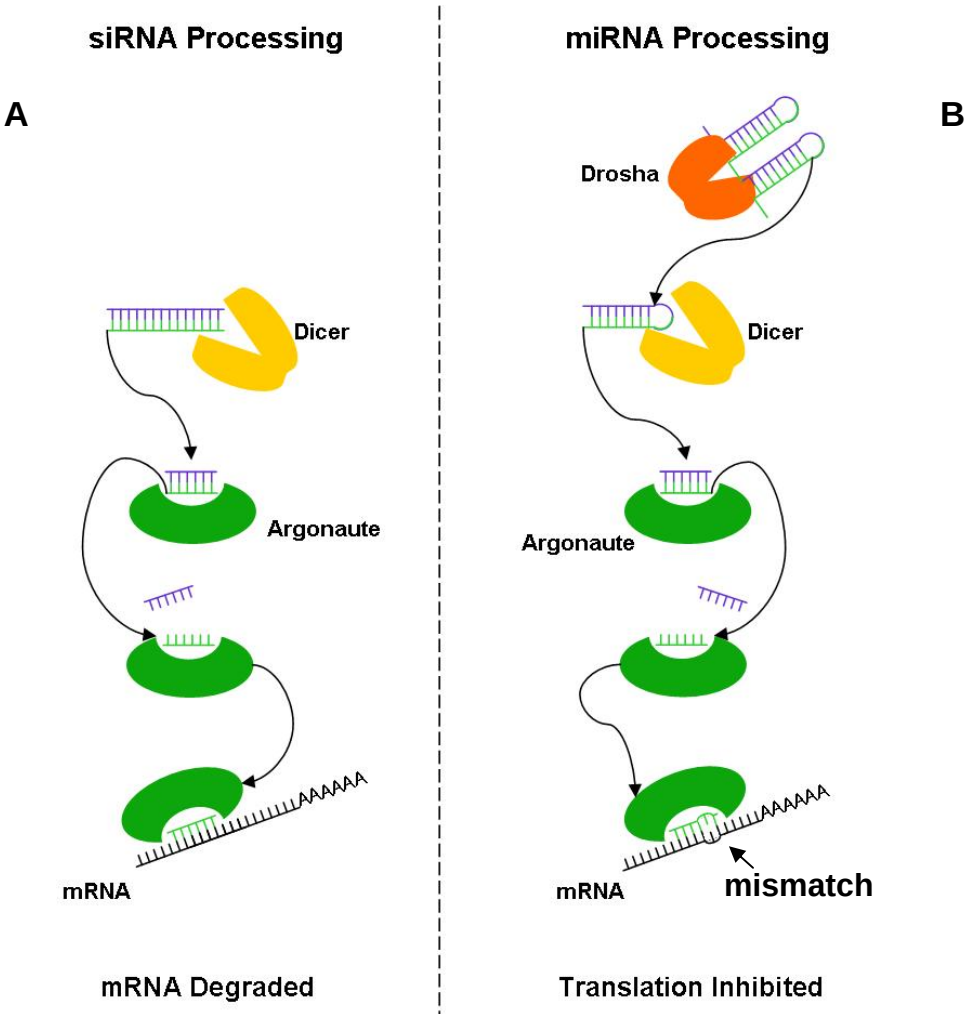


Figure 1.6 – The processing of small regulatory RNAs from longer precursor RNA required for post-transcriptional gene silencing (PTGS).

(A) Short interfering (si)RNA is produced from long, linear, double-stranded RNA precursors which are diced into siRNA by the RNaseIII activity of Dicer. The siRNA duplexes are taken into the Argonaute protein before the passenger strand is released, leaving the guide strand bound to Argonaute. The Argonaute then binds other proteins to form the RNA-induced silencing complex (RISC) and the guide strand targets the complex to homologous mRNA which is sliced by Argonaute and thus degraded.

(B) Micro (mi)RNA is produced from long hairpin RNA precursors which are initially cleaved by the RNaseIII enzyme Drosha before being further cut down by Dicer's RNaseIII activity. As with siRNA, the resulting miRNA duplexes are taken into Argonaute where the passenger strand is released, leaving the mature guide strand bound to Argonaute. The Argonaute is then assembled into the miRNA-containing ribonucleoprotein particle (miRNP) complex which is targeted to mRNA that is homologous to the guide strand. miRNA is different to siRNA because it contains mismatches to the mRNA target which results in inhibition of translation rather than mRNA degradation.

1.3.3 RNAi in heterochromatin formation

1.3.3.1 The RNAi-mediated heterochromatin formation pathway

In the fission yeast *S. pombe*, the main role of RNAi seems to be in silencing of constitutive heterochromatin loci through Transcriptional Gene Silencing (TGS), associated with formation of heterochromatin, as well as transcript degradation via cis-acting post-transcriptional gene silencing (PTGS). There are three main protein complexes involved in TGS in *S. pombe*: (1) *R*NA-induced *I*nitiation of *T*ranscriptional gene *S*ilencing (RITS) Complex (Verdel, 2004) (2) RNA-dependent RNA Polymerase Complex (RdRC) (Motamedi et al., 2004) (3) Clr4 Complex (ClrC) (Horn et al., 2005).

The **R**NA-induced **I**nitiation of **T**ranscriptional gene Silencing (**RITS**) Complex	
Name	**Function**
Tas3	Argonaute Binding Protein
Chp1	Binds H3K9me2 via Chromodomain
Ago1	Binds small RNA, Argonaute Family Protein
The **Clr**4 **C**omplex (**ClrC**)	
Name	**Function**
Clr4	H3K9-specific Methyltransferase
Rik1	WD-repeat containing protein
Dos1^{Raf1}	WD-repeat containing protein
Dos2^{Raf2}	Unknown Function
Cul4	Ubiquitin Ligase Scaffold Protein
Ned8	Ubiquitin-like protein
The **R**NA-directed **R**NA Polymerase **C**omplex (**RdRC**)	
Name	**Function**
Rdp1	RNA-dependent RNA Polymerase
Hrr1	RNA Helicase
Cid12	Putative Poly-A Polymerase

Table 1.1 – A summary of the three main protein complexes involved in the targeting of heterochromatin formation by the RNAi pathway in *S. pombe*.

RITS is the main RNAi effector complex in this system and it contains the Argonaute protein Ago1 which is bound to the siRNA guide strand. The RdRC amplifies the siRNA levels whilst the ClrC methylates lysine9 of Histone H3 (H3K9me) causing the formation of heterochromatin. An interesting characteristic of this pathway is the fact that it forms a self-reinforcing loop (Figure 1.7), thus removing a single component can cause the entire system to malfunction (Noma et al., 2004).

The dg/dh repeats of the pericentromere are transcribed by RNA Polymerase II (RNAPII) and the transcripts are used to produce siRNAs (Djupedal, 2005). The RNA-dependent RNA polymerase activity of the RdRC is also required, however, in order to amplify the level of siRNAs in the cell for functional pericentromeric silencing (Sugiyama et al., 2005), presumably using the RNA transcript as a template. The next step in siRNA generation is dicing of the transcript into ~25nt duplex siRNAs by the RNaseIII enzyme Dicer (*Dcr1*) (Volpe et al., 2002).

The duplex siRNA is now loaded into the sole Argonaute protein of *S. pombe*, Ago1. The mechanism of loading in *S. pombe* is unknown, although the Heat-shock proteins Hsp70 and Hsp90 have been implicated in Ago1 loading in other organisms (Iwasaki et al., 2010), (Iki et al., 2010), (Miyoshi et al., 2010). At this point Ago1 is complexed with Arb1 and Arb2 in the *A*rgonaute si*R*NA *c*haperone (ARC) complex, which acts to prevent maturation of the duplex siRNA into the fully functional single stranded siRNA (Buker et al., 2007).

The 'slicer' activity of Ago1 is required for efficient maturation of the siRNA duplex, removing the passenger strand whilst keeping the guide strand bound (Buker et al., 2007). The mature guide strand-loaded Ago1 can now associate with Tas3 and Chp1 to form the RITS complex (Verdel, 2004). The complex is localised to heterochromatin through the binding of Chp1 to H3K9me via its chromodomain (Partridge et al., 2002) and the targeting of the Ago1-bound siRNA to the homologous centromeric transcript (Partridge et al., 2007).

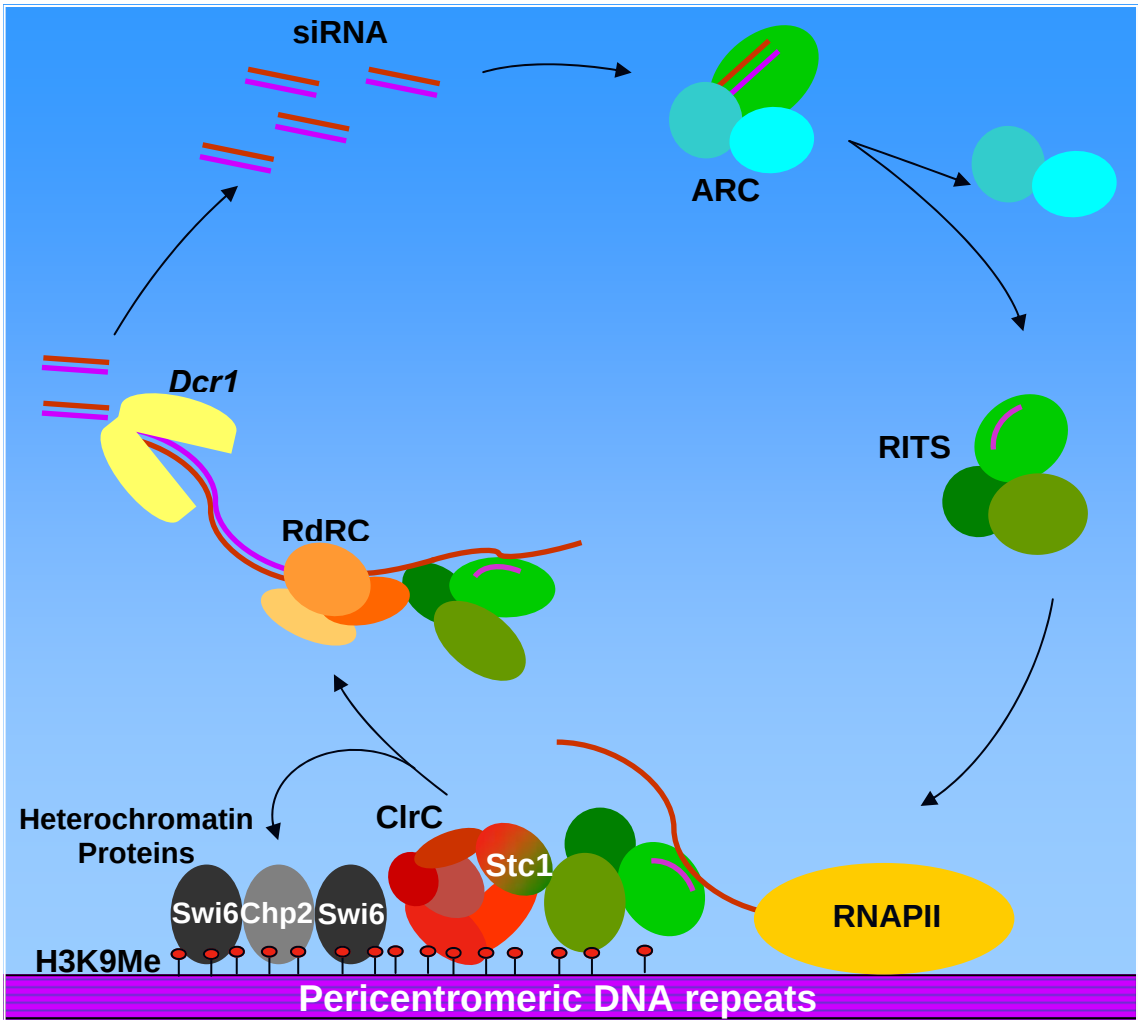


Figure 1.7 – RNAi-mediated heterochromatin formation in *S. pombe*. Transcription of the pericentromeric repeat region is required in order to make short interfering (si)RNAs which feedback to cause transcriptional silencing of the repeats.

Heterochromatin formation requires the recruitment of the H3K9 histone methyltransferase (HMT) Clr4. RITS can recruit Clr4 to the pericentromere via its interaction with a recently characterised bridging protein, Stc1 (Bayne et al., 2010). In their investigation it was found that tethering of Stc1 to a euchromatic locus is enough to form RNAi-independent heterochromatin. H3K9 methylation acts as a platform to recruit chromodomain-containing proteins such as: the heterochromatin proteins Swi6 and Chp2, the RITS complex protein Chp1, the HMT Clr4, the Clr6 HDAC Complex protein Alp13. Recruitment of the heterochromatin proteins is particularly important for formation of the condensed structure of heterochromatin and Swi6 and Chp2 have distinct roles in this process (Motamedi et al., 2008), (Fischer et al., 2009).

Despite the fact we know that Clr4 is the sole histone methyltransferase that methylates histone H3 to produce H3K9 methylation in *S. pombe* (Nakayama et al., 2001), the role of the ClrC as a whole in the RNAi-dependent heterochromatin formation pathway is surprisingly enigmatic. The ClrC complex displays E3 ubiquitin ligase activity *in vitro* and the full complex is required for heterochromatin formation (Horn et al., 2005), (Ekwall and Ruusala, 1994). No substrates have yet been identified for this ubiquitin ligase activity *in vivo*.

Interestingly a new non-histone substrate for the methyltransferase activity of Clr4 was recently identified, the nuclear RNA export factor Mlo3 (Zhang et al., 2011). In this study Co-IP experiments were used to show that Mlo3 interacts with the ClrC, TRAMP subunits and also with the RITS protein Chp1 in a Clr4-dependent manner *in vivo*. *mlo3*Δ was shown to have defective siRNA processing whilst H3K9 methylation and Swi6 at the heterochromatin loci remained intact. Together this data suggests that Clr4 may regulate RNAi through methylation of Mlo3 which is important for siRNA processing, perhaps by identifying aberrant RNA for degradation by the exosome.

Two interesting questions regarding the initiation of the RNAi-mediated heterochromatin formation loop that remain to be answered are; (1) How is the positive feedback loop of the RNAi-mediated heterochromatin formation pathway initiated? (2) How are double-stranded (duplex) siRNAs produced from

a single-stranded transcript? One theory to explain both questions is that the single-stranded transcripts form a double-stranded hairpin structure which allows primary duplex siRNAs to be produced independently of RdRC which kickstart the pathway before the RdRC amplifies the siRNA levels (Djupedal et al., 2009). An alternative theory for the initiation of this pathway is that primal (pri)RNAs generated by centromeric RNA degradation products are taken up by Ago1 which can then recruit the RdRC to the pericentromeric repeats (Halic and Moazed, 2010). Another explanation for the appearance of duplex siRNA is that the RdRC, whilst using the transcript as a template for the amplification of siRNA, is simultaneously producing double-stranded RNA.

1.3.3.2 The role of HDACs and histone remodeling enzymes in heterochromatin formation

Heterochromatin formation at all of the constitutive heterochromatin loci is also dependent on a number of other histone modifying and remodeling enzymes. Histone deacetylase enzymes (HDACs) allow spreading of heterochromatin by removing the chromatin acetylation marks associated with euchromatin. There are three HDACs involved in deacetylation of histones in heterochromatin regions; Clr3, Clr6 and Sir2 (Grewal et al., 1998), (Wiren et al., 2005).

Sir2 has a limited role in heterochromatin formation at the centromere but is essential for forming heterochromatin at the mating-type and subtelomere (Shankaranarayana et al., 2003). At the mating-type and subtelomeres Sir2 works upstream of Swi6 recruitment to heterochromatin suggesting an important role in establishing heterochromatin at these loci (Freeman-Cook et al., 2005).

The Clr3-containing SHREC (Snf2/HDAC-containing Repressor Complex) complex is localised to all of the heterochromatin loci (Sugiyama et al., 2007) and it can be recruited to heterochromatin through interactions with the heterochromatin proteins Chp2 and Swi6 or via heterochromatin nucleation centres through RNAi and DNA-binding proteins (Yamada et al., 2005), (Fischer et al., 2009). In addition to HDAC activity of SHREC it also harbors

chromatin remodeling Snf2 ATPase activity and both activities are essential for restricting the occupancy of RNAPII at the pericentromeres (Sugiyama et al., 2007). Deletion of a SHREC subunit gene causes a loss of silencing at the pericentromere but interestingly there is also an increase in both siRNA and repeat transcript levels suggesting that loss of SHREC results in a more active recruitment of RNAPII transcription compared to simply losing heterochromatin (Sugiyama et al., 2007).

Clr6 has quite a promiscuous HDAC activity and mutants of this essential protein display general hyperacetylation of the genome (Wiren et al., 2005). Clr6 Complex II has been shown to interact with Swi6 and the HDAC appears to have an overlapping role in transcriptional repression with Clr3 at the centromere, although Clr6 has a much smaller affect on heterochromatin than Clr3 (Fischer et al., 2009), (Bjerling et al., 2002).

The ubiquitination of H2B has also been found to regulate heterochromatin. The HULC (histone H2B ubiquitin ligase complex) complex ubiquitinates H2B on K119 (Zofall and Grewal, 2007). Deletion of HULC subunits results in enhanced silencing of marker genes at the pericentromere whilst overexpression of the Rhp6 subunit results in silencing defects and increased RNAPII occupancy (Zofall and Grewal, 2007).

The histone regulatory complex A (HIRA) complex is a histone chaperone known to affect heterochromatic silencing and chromosome segregation suggesting a role in heterochromatin formation. The complex is made up of four proteins; Slm9, Hip1, Hip3, Hip4 (Blackwell et al., 2004), (Greenall et al., 2006), (Anderson et al., 2010). Deletion of any of the four proteins causes alleviation of silencing of marker genes at the pericentromeric and mating-type heterochromatin loci, as well as loss of silencing of Tf2 retrotransposons (Blackwell et al., 2004), (Anderson et al., 2010). Deletion of the H3-H4 chaperone $Cia1^{Asf1}$ causes the same set of defective silencing phenotypes as HIRA-subunit null strains and Cia1 interacts with HIRA proteins, suggesting that the chaperones act in concert (Yamane et al., 2011). Localisation of HIRA-Cia1 across the three constitutive heterochromatin loci is Swi6-dependent and Hip1 has been shown to interact with Swi6 (Yamane et al., 2011). HIRA-Cia1

interacts with Clr6 Complex II and is required for H3K9 deacetylation (Yamane et al., 2011).

1.3.4 Initiation of pericentromeric heterochromatin formation

The first step in initiation of heterochromatin formation at the centromere is difficult to determine due to the seeming codependence of H3K9me and RNAi. A number of experiments have provided insights into the order of H3K9 methylation and RNAi localisation in the formation of heterochromatin at this locus.

A series of genetic experiments were used to test the effects of deletion of the chromodomain-containing RITS protein Chp1 and the H3K9me HMT enzyme Clr4 on each of the three main heterochromatin loci in fission yeast (Sadaie et al., 2004). They found that deletion of Chp1 resulted in reduced levels of H3K9me only at the centromere whereas deletion of Clr4 resulted in loss of H3K9me at all loci. Upon deletion of Clr4 and consequent reintroduction of the enzyme in a *chp1*Δ mutant no H3K9me could be established at any of the three loci suggesting that RITS is required for establishment but not maintenance of heterochromatin in *S. pombe*. This result is the basis for a model of heterochromatin formation in which the recruitment of RITS and RNAi precedes H3K9 methylation.

An experiment tethering the RITS subunit Tas3 to the transcripts of a *ura4*+ marker gene provides support for the model in which RNAi precedes H3K9 methylation (Buhler et al., 2006). The marker gene was silenced via the tether in a manner dependent on RNAi, ClrC, HDACs and heterochromatin proteins, suggesting that RITS can recruit all of the factors required for heterochromatin formation – in the absence of initial H3K9me - via its interaction with transcripts. Tethering Rdp1 to the transcript did not result in silencing, suggesting that the RITS complex's role is more intricate than simply recruiting the machinery for siRNA amplification to the locus. The fact that this artificially induced heterochromatin is still dependent on siRNA production

correlates well with the finding that spreading of heterochromatin at the pericentromere is dependent on the slicer activity of Ago1 (Irvine et al., 2006).

Experiments using mutations in Tas3, the RITS component responsible for bridging Chp1 and Ago1, have been used to elucidate the chronology of events leading to RNAi-mediated heterochromatin formation (Partridge et al., 2007), (DeBeauchamp et al., 2008). The $tas3_{WG}$ mutant loses the Tas3-Ago1 interaction but exhibits only very weak heterochromatin formation defects (Partridge et al., 2007). Through transient depletion of H3K9me it was found that the $tas3_{WG}$ mutant can maintain centromeric heterochromatin but not establish it. In another study the Tas3-Chp1 interaction was lost in the $tas3_{\Delta 10\text{-}24}$ mutant resulting in a strong defect in heterochromatin formation, suggesting that Ago1 is dependent on the Chp1-Tas3 interaction for recruitment to the centromere (DeBeauchamp et al., 2008). As deletion of ClrC components results in complete loss of H3K9 methylation at all heterochromatin loci whereas deletion of RNAi components results only in a lower level of H3K9 methylation at the centromere, the authors of the two experiments use their results to propose a new model for establishment of RNAi-mediated heterochromatin at centromeres. In this model an RNAi-independent mechanism initially recruits Clr4 to centromeres to produce low levels of H3K9me, RITS can then be recruited through the chromodomain of Chp1 allowing the consequent recruitment of the RdRC for siRNA amplification which feeds back to produce robust heterochromatin formation. An alternative interpretation of the data from both experiments is that RITS is recruited to the centromere in the absence of H3K9 methylation through the siRNA guide associated with Ago1, allowing Clr4 recruitment to produce H3K9 methylation which stabilises RITS localisation through the Chp1 chromodomain. The $tas3_{WG}$ can maintain heterochromatin because both Chp1-Tas3 and Ago1 are recruited to the centromere through the Chp1 chromodomain and via the siRNA guide strand respectively whereas in $tas3_{\Delta 10\text{-}24}$ centromeric heterochromatin may be lost because the Chp1-Tas3 interaction may be important for recruitment of the RdRC and/or Dicer.

Another artificial tethering system has provided additional insights by further dissecting the roles of factors in the formation of heterochromatin. Clr4 was artificially tethered to DNA at an *ade6* marker gene by a GBD binding system (Kagansky et al., 2009). In this system the silencing of the marker gene is RNAi-independent but requires ClrC components, HDACs and heterochromatin proteins. This result suggests that the main role of RNAi in de novo heterochromatin formation is as a means of recruiting ClrC in the absence of H3K9me. The Clr4 tether system has also been used to distinguish between the requirement of proteins for establishment of heterochromatin and maintenance of heterochromatin. In this assay RNAi proteins and Swi6 are not required for establishment or maintenance of heterochromatin whereas ClrC, the HDACs Sir2 and Clr3 and the heterochromatin protein Chp2 are required for both. Interestingly when *stc1*Δ is tested in the Clr4 tether system it has a phenotype part way between ClrC and RNAi proteins, in that Stc1 is required for establishment but not for maintenance of heterochromatin (Bayne et al., 2010).

1.3.5 The role of essential proteins in pericentromeric heterochromatin

Several genetic screens have been carried out in order to identify novel factors involved in gene silencing in *S. pombe*. In 1994 a screen for loss of silencing at the mating type locus identified Clr3, Clr4 and Rik1 as proteins required for silencing (Ekwall and Ruusala, 1994). A similar screen was carried out by Ekwall et al some years later in order to identify novel proteins required for silencing at the pericentromere, but not the mating-type, and 12 unique loci were found to harbour mutations causing defective silencing. This mutant set was named the *C*entromere *S*uppressor of *P*osition Effect (*csp*) mutants (Ekwall, 1999). The current status of the *csp* screen is summarised in Table 1.2.

Several *csp* mutants were found to be alleles of core RNAi proteins which had been identified by other labs as important for heterochromatin formation at the centromere. As the RNAi pathway is dispensable for survival in *S. pombe* none of the *csp* RNAi mutants were temperature sensitive (ts). In addition several

ts mutants were found – those that have been identified so far are all proteins that are required in other essential pathways but also have a role in RNAi. *csp3* was identified as an allele of the RNA Polymerase II (RNAPII) subunit rpb7 and it was shown that the pre-siRNA transcripts are produced by RNAPII (Djupedal, 2005). More recently it was shown that *csp4* and *csp5* are splicing factors and that a number of other splicing factors are also directly involved in RNAi (Bayne et al., 2008).

ts Mutants		
Mutant	**Gene Identified**	**Role in *S. pombe***
csp1	---------?---------	
csp2	---------?---------	
csp3	rpb7	RNAPII Subunit
csp4	cwf10	Splicing Factor
csp5	prp39	Splicing Factor
csp6*	***ssa2***	**Hsp70 Protein**
non-ts Mutants		
Mutant	**Gene Identified**	**Role in *S. pombe***
csp7	rdp1	RNAi Component
csp8	cid12	RNAi Component
csp9	ago1	RNAi Component
csp10	cid12	RNAi Component
csp11	---------?---------	
csp12	arb1	RNAi Component
csp13	---------?---------	

* *csp6* was identified during the course of this study

Table 1.2 – A summary of the mutants found during the *C*entromere *S*uppressor of *P*osition Effect (*csp*) mutant screen (Ekwall 1999).

1.3.5.1 Heat-shock proteins and their role in gene silencing

The heat shock protein (Hsp)70 family of proteins are named for their essential functions as chaperones which protect the cell's proteins from denaturation whilst the cell endures high temperatures, along with other cell stresses. This protein family is very highly conserved from bacteria to mammals. In most organisms there are a number of both constitutively expressed and stress-induced Hsp70s as well as Hsp70s with localisation in specific cellular compartments. In humans, for example, there are 11 Hsp70 proteins including the constitutively expressed Hsc70^{HSPA8}, the stress-induced Hsp70 proteins HSPA1A, HSPA1B, HSPA1L, the endoplasmic reticulum (ER)-localised BiP^{HSPA5} and the mitochondrion-localised mtHsp70^{HSPA9} (Tavaria et al., 1996). These proteins carry out a variety of functions and roles including protein folding/refolding (Hartl and Hayer-Hartl, 2002), translocation of proteins across membranes (Tomkiewicz et al., 2007), regulation of signaling pathways (Pratt and Toft, 2003) and regulation of protein degradation (Luo et al., 2009).

Recently Hsp70 has been implicated in the RNAi pathway in numerous independent investigations. Three *in vitro* studies using plant or fly extracts showed that Hsp70 and Hsp90 are required for the efficient loading of siRNAs into Argonaute (Iki et al., 2010), (Iwasaki et al., 2010), (Miyoshi et al., 2010). Inhibition of Hsp90 has also been shown to result in a decrease in Argonaute protein levels in human cells suggesting that Hsp90 interactions stabilise Argonautes (Johnston et al., 2010).

Hsp70 proteins have a highly conserved tertiary structure made up of two domains; the N-terminal Nucleotide Binding Domain (NBD), which has ATPase activity, and the C-terminal Substrate Binding Domain (SBD), which binds to the protein substrate (Jiang et al., 2006). The SBD has two sub-domains consisting of a binding pocket which can bind five consecutive hydrophobic residues of substrate proteins (Rudiger et al., 1997), and also an α-helical bundle which acts as a lid to open/close over the binding site (Zhu et al., 1996). In the ATP-bound state the 'lid' is open allowing quick substrate binding and release, whereas upon hydrolysis of ATP to ADP a conformational change is transmitted

from the NBD to the lid domain causing it to close over the substrate binding site and more stably capture the substrate (Zhu et al., 1996).

In order to fulfill their numerous roles and functions in the cell, the Hsp70 proteins interact with various cofactors which can regulate their localisation as well as their substrate specificity and binding capabilities. Hsp70 activity is mainly regulated is through its nucleotide cycle of ATP binding, ATP hydrolysis to ADP, and exchange of ADP for ATP (Figure 1.8). The ATPase activity of Hsp70 is quite weak, however it is mildly stimulated by substrate binding, and stimulated to a greater extent by cofactors (Liberek et al., 1991), predominantly the J proteins, causing Hsp70 substrate binding and release to slow down and become more stable. Nucleotide Exchange Factors (NEFs) are required in order to recycle the Hsp70 protein from its ADP-bound state, in which the substrate is slow to be released, to its ATP-bound state which has a fast substrate release (Szabo et al., 1994).

NEFs are an evolutionarily unrelated group of proteins which have convergent functions in nucleotide exchange, although the exchange mechanisms differ from protein to protein (Cyr, 2008). An example of a NEF is the Hsp110 protein which is part of the Hsp70 superfamily and requires binding of a nucleotide to its NBD in order to switch on NEF activity.

J proteins are a highly diverse family of proteins which share a common J domain. The J domain interacts with Hsp70 via its NBD and its linker region, which resides between the NBD and the SBD, to stimulate ATP hydrolysis (Jiang et al., 2007). Interestingly there are many more different J proteins in the cell than Hsp70 paralogues and it has been hypothesised that the interaction between a specific Hsp70 and its subset of J proteins is a major determinant for the diverse non-redundant roles of the cell's multiple Hsp70s (Kampinga and Craig, 2010). An example of this is the case of two human J proteins which both interact with the Hsp70 protein HSPA1A. The J protein HSPB1 is known to help HSPA1A to refold proteins after heat shock whereas the J protein HSPB2, which contains ubiquitin interacting motifs, can also interact with HSPA1A in order to redirect the denatured proteins to be degraded instead of refolded (Kampinga and Craig, 2010).

The J domain is used as a means to recruit Hsp70 and it is found in proteins which are otherwise totally dissimilar to one another in terms of their domain structure. Aside from the vital role of the J domain in stimulating the ATPase activity of Hsp70, the other domains of J proteins can affect their function. J proteins with a transmembrane domain can be tethered to a particular membrane in a site that has a high concentration of unfolded proteins such as the ribosome in order to recruit Hsp70 (Otto et al., 2005). Some J proteins contain a substrate binding domain and can bind to substrate proteins independently of Hsp70 before passing the substrate to Hsp70 (Shen and Hendershot, 2005).

Hsp70 and its cofactors have been shown to team up with Hsp90 which is a chaperone responsible for the stability of over 200 specific client proteins in humans, many of which are known oncogenic proteins (Picard, 2011). The best characterised example of this Hsp70/Hsp90 system is in the regulation of steroid hormone receptors (SRs). Initially Hsp40, a J protein, binds the SR and recruits Hsp70 which then binds the SR (Daniel et al., 2007). Hsp70 stabilises the client protein and may partially open its conformation, before passing the client protein to Hsp90 which opens the steroid binding site before finally dissociating upon steroid binding (Pratt and Toft, 2003). Hsp70/Hsp90 organising protein (Hop) binds to Hsp70 and recruits Hsp90 in this system, and numerous Hsp70 and Hsp90 cofactors are also required.

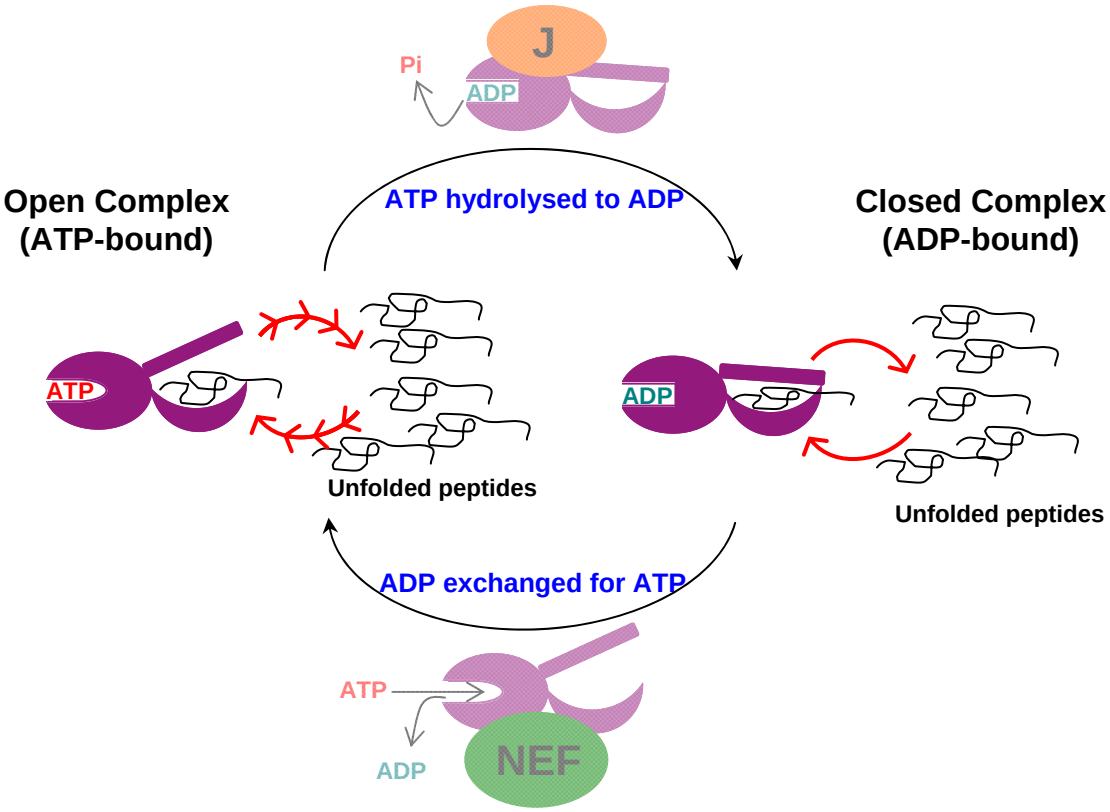


Figure 1.8 – The nucleotide binding cycle of Hsp70.
The nucleotide bound in the nucleotide binding domain (NBD) of Hsp70 affects the conformation of the protein, thus allowing regulation of its activity. When ATP is bound to the NBD then the substrate binding domain (SBD) is in an open conformation and there is fast binding and release of substrate peptides. When ADP is bound at the NBD then the SBD is in a closed conformation where the substrate is stably captured by the chaperone.
The correct balance between substrate binding/release and stable substrate capture is required for optimum substrate (re)folding thus correct regulation of the ATP and ADP bound states of Hsp70 is vital.
Substrate binding and J proteins stimulate the ATPase activity of the NBD causing hydrolysis of ATP to ADP. Nucleotide exchange factors (NEFs) stimulate release of ADP from the NBD allowing consequent binding of ATP.

1.3.5.2 RNA Polymerase II

The RNA Polymerase II (RNAPII) complex is a highly conserved member of the DNA-dependent RNA Polymerase family. This family is responsible for the transcription of a DNA template into RNA copies and, aside from plants, there are normally three RNA Polymerase complexes in eukaryotes each of which are responsible for transcribing particular subsets of the genome. RNAPII transcribes the protein coding genes into messenger (m)RNA and more recently it has been found to transcribe the bulk of ncRNAs in the genome, including the precursor transcript for siRNA generation in *S. pombe* (Djupedal, 2005).

The RNAPII complex is made up of 12 subunits, all of which are essential. The subunits are numbered in relation to their size starting with the largest subunit, therefore in *S. pombe* they are Rpb1 through to Rpb12 (Figure 1.9). Studies in bacteria found that a subassembly complex is formed before the full RNAP machinery is matured (Ishihama, 1981). This complex was found to be made up of 2 α subunits along with the β subunit. In eukaryotes the Rpb2 subunit is homologous to the β subunit, whereas the two α subunits have diverged and are represented by the Rpb3 and Rpb11 subunits (Kimura et al., 1997). The Rpb3 subunit in yeast, similarly to the bacterial α subunit (Gourse et al., 2000), has been linked to promoter recognition (De Angelis et al., 2003), Soutourina 2011 (Soutourina et al., 2011). In mammals it has also been shown that Rpb3 has retained a key role in RNAPII assembly forming strong interactions with all of the subunits except Rpb6, Rpb8 and Rpb9 (Acker et al., 1997).

The C-terminal Domain (CTD) of Rpb1 is made up of a serine-rich heptad repeat which is variable in length across species but is essential for cell viability. Phosphorylation of Ser2 and Ser5 of the CTD are important in transcriptional initiation, elongation and for co-transcriptional RNA processing (Buratowski, 2009). Splicing of mRNA is carried out co-transcriptionally, and the CTD acts as a 'landing pad' for splicing factors during transcriptional elongation (Zorio and Bentley, 2004). The speed of transcription can affect the way in which the transcript is spliced (de la Mata et al., 2003) and also splicing

factors can affect the speed of transcription (Fong and Zhou, 2001). Splicing factors were recently found to be required for correct function of RNAi, with splicing factor mutants exhibiting defects in heterochromatin integrity even under conditions in which splicing was unaffected (Bayne et al., 2008).

Due to the crucial role of RNAPII in the cell there are a whole host of proteins which aid and regulate its activity, including general transcription factors, chromatin remodeling enzymes, activators, coactivators and repressors. Before the RNAPII complex is recruited to a promoter, transcription factors (TFs) initially destabilise nucleosomes around the promoter region, and the surrounding histones are acetylated by histone acetyltransferases (HATs) (Li et al., 2007). The pre-initiation complex (PIC), which is composed of RNAPII (with a hypophosphorylated CTD) along with the general transcription factors (GTFs), is then recruited to the site via TFs and coactivators such as the Mediator complex (Li et al., 2007). During initiation of transcription the CTD of Rpb1 is phosphorylated on Ser5 and as the RNAPII complex gets further from the promoter during elongation there is a gradual decrease in Ser5 phosphorylation accompanied by an increase in Ser2 phosphorylation. The changes in Rpb1 CTD phosphorylation result in many of the GTFs to be swapped out for elongation factors and mRNA processing factors (Hahn, 2004).

When RNAPII transcribes a region of the genome the DNA must be unwound, and the histones have to be moved/removed and consequently replaced upstream of the elongating polymerase (Selth et al., 2010). The histones carry important epigenetic information concerning the level of transcription that the DNA should undergo as well as the type of chromatin that the DNA region should be packaged in (Jenuwein and Allis, 2001). Transcription may act as a chance for the cell to reassess a chromatin state as some of the histones are evicted and replaced during elongation (Workman, 2006).

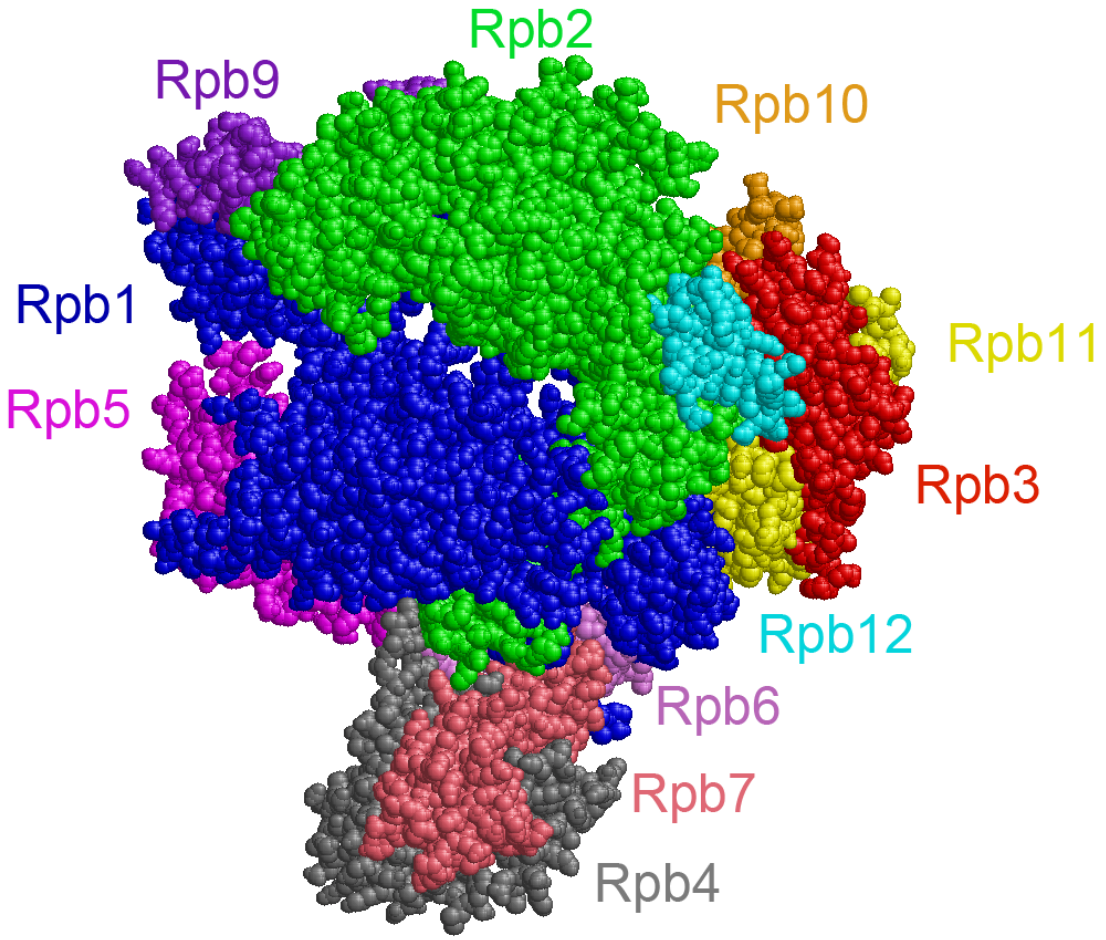


Figure 1.9 – The RNA Polymerase II complex. The complex is made up of 12 subunits. Rpb8 is not visible in this figure as it is situated on the hidden side of the enzyme in a similar position to Rpb12.
The crystal structure used is 1WCM (Armache, 2005)

In mammals it has been found that ncRNAs can regulate RNAPII activity (Ponicsan et al., 2010). The best characterised examples of these ncRNAs are the B2 RNAs in mice (Allen et al., 2004) and the Alu RNAs in humans (Mariner et al., 2008). These ncRNAs are assembled into the RNAPII pre-initiation complex and act during heat shock to repress normally highly expressed genes such as actin (Espinoza et al., 2004), (Mariner et al., 2008). Varying lengths of synthetic poly-guanosine oligonucleotide sequences were used to show that there is an RNA docking site on RNAPII (Kugel and Goodrich, 2002) which was later found to be used by the B2 RNA (Espinoza et al., 2004), although the exact position of this docking site is unknown.

1.3.5.3 The role of RNAPII at centromeres

RNAPII transcribes the pericentromeric (dg/dh) repeats in *S. pombe*. Transcription in this region has directional regulation, because the reverse strand is constitutively expressed whereas the forward strand is repressed by heterochromatin (Volpe et al., 2002). More recent analysis of repeat transcription during the cell cycle has shown that both forward and reverse transcript levels peak during S phase when the pericentromeric locus is undergoing replication (Kloc et al., 2008). This cell cycle regulation of repeat transcript levels suggests that RITS may be mainly loaded with siRNA after S phase, perhaps in readiness to spread H3K9me onto the newly loaded histones in the pericentromeric locus.

The temperature sensitive (ts) RNAPII mutant *rpb7-G150D*^{sp3} displays a defect in heterochromatin formation at the pericentromeric repeats. In this mutant, general transcription is mildly affected at the permissive temperature, and all but lost at the restrictive temperature (Djupedal, 2005). It appears that there is a specific, stronger defect in transcription of the reverse strand of the dg/dh repeats even at the permissive temperature, suggesting that the Rpb7 subunit is particularly important for transcription of the pre-siRNA transcript. The Rpb7 subunit is known to be particularly important for initiation of transcription, along with Rpb4, and the two subunits form a close binding pair which is dissociable from the rest of RNAPII in budding yeast (Armache et al., 2003). Rpb7 is likely to interact with the exiting RNA chain and it has been suggested that Rpb7 presents a good interaction surface for regulatory proteins due to its large exposed surface area and its orientation towards the DNA upstream of RNAPII during transcription (Armache et al., 2003).

A non-transcriptional role for RNAPII in the RNAi pathway has been inferred by the analysis of the *rpb2-m203* mutant (Kato, 2005). Unlike *rpb7-G150D*, *rpb2-m203* does not have any obvious growth defects and transcription of the dg/dh repeats is intact. *rpb2-m203* has typical RNAi pathway defects, with reduction of the heterochromatin mark H3K9me2 specifically at the pericentromere, decreased levels of siRNA and accumulation of pre-siRNA transcript. The mechanism by which a mutation in Rpb2 causes this RNAi

defect still remains to be elucidated. One possibility is that the processing of the pre-siRNA transcript is co-transcriptional, in a similar way to splicing/capping of mRNA where the RNAPII complex acts as a platform for recruiting processing factors (Zorio and Bentley, 2004). The residue mutated in *rpb2-m203* is exposed on the surface of the RNAPII complex and so it is possible that this surface is important for RNAi.

Recently it has been discovered that the central core region of the *S. pombe* centromeres is also transcribed (Choi et al., 2011). In this study it was shown that these transcripts under the kinetochore (TUKs) accumulate, with varying lengths, in exosome mutants but not in wildtype cells suggesting that they are rapidly degraded by the exosome. Mutants with reduced levels of CENP-A at the centromere accumulated TUKs displaying discrete lengths. The TUKs were shown to be initiated from within the central core, ruling out transcriptional run-on from the outer repeats. It is likely that these TUKs are transcribed by RNAPII as they are polyadenylated and have a 5' cap. It is currently unknown if TUKs provide a function at the centromere.

1.4 Summary

Much has been learned about the non-essential core components involved in RNAi-mediated heterochromatin formation in *S. pombe* in the last decade. The *csp* screen revealed that essential processes, such as transcription, clearly contribute to heterochromatin formation. In order to learn more about the role of RNAPII in RNAi-mediated heterochromatin formation I carried out a screen for mutations in the RNAPII subunits Rpb3 and Rpb11 that cause defects in heterochromatin formation. Characterisation of the resulting mutants has led to some insight into the complex role that RNAPII plays in this pathway. I also identified the affected gene in the *csp6* mutants to be a Hsp70 gene and I have explored its possible role in loading of siRNA into the RITS complex.

Chapter 2 - Materials and Methods

2.1 *S. pombe* culture and media

Growth

S. pombe cultures and colonies were incubated at temperatures between 18ºC and 36ºC for between overnight and 3 days as indicated for each experiment. Wildtype haploid strains will grow with the following generation times:

Medium	Temperature °C	Generation Time
YES	25	3 hours
	32	2 hours 10 minutes
	36	2 hours
PMG (minimal)	25	4 hours
	32	2 hours 30 minutes
	36	2 hours 20 minutes

The generation time of mutant strains may vary. The strains with mutated or deleted Hsp70 or RNAPII subunit genes all grow more slowly than wildtype, with greater than doubled generation times for the worst affected strains. The time required to double the population of cells can be accurately calculated using the following formula: $T= \log(2^{t2-t1})/\log(x/y)$, where T is the generation time, x is cells per ml at t1 and y is cells per ml at t2.

2.1.1 Growth Media

All solutions were made up to a final volume in dH_2O and autoclaved unless otherwise stated.

PMG agar in 900ml:

Pthallic acid	3g
di-sodium orthophosphate	2.2g

glutamic acid	3.75g
D-glucose anhydrous	20g
vitamins 1000x	1ml
minerals 10,000x	0.1ml
salts 50x	20ml
agar (OXOID)	20g

PMG liquid in 900ml:

Pthallic acid	3g
di-sodium orthophosphate	2.2g
glutamic acid	3.75g
D-glucose anhydrous	20g
vitamins 1000x	1ml
minerals 10,000x	0.1ml
salts 50x	20ml

YES agar (-ade):

Yeast extract (DIFCO)	5g
D-glucose anhydrous	30g
Arginine (Sigma)	0.2g
Lysine (Sigma)	0.2g
Histidine (Sigma)	0.2g
Uracil (Sigma)	0.2g
Leucine (Sigma)	0.2g
Agar (OXOID)	20g

YES liquid:

Yeast extract (DIFCO)	5g
D-glucose anhydrous	30g
Arginine (Sigma)	0.2g

	Lysine (Sigma)	0.2g
	Histidine (Sigma)	0.2g
	Uracil (Sigma)	0.2g
	Leucine (Sigma)	0.2g

4 x YES liquid: As above all reagents x 4.

ME plates (1L):

	Malt extract (OXOID)	30g/L
	Adenine (Sigma)	250g/L
	Arginine (Sigma)	250g/L
	Histidine (Sigma)	250g/L
	Uracil (Sigma)	250g/L
	Leucine (Sigma)	250g/L

Vitamins 1000x (100ml):

	Pantothenic acid	0.5g
	Nicotinic acid	1g
	Inositol	1g
	Biotin	1mg
	Filter sterilised	

Minerals 10,000x (100ml):

	Boric acid	5g
	$MnSO_4$	4g
	$ZnSO_4$	4g
	$FeCl_26H_2O$	2g
	Molybdic acid	1.6g
	$CuSO_4$ $5H_2O$	0.4g
	Citric acid	10g
	Filter sterilised	

Salts 50x:

Magnesium chloride	53.5g
Calcium chloride	1g
Potassium chloride	50g
di-sodium sulphate	2g

Supplement stocks:

Adenine 50x (Sigma)	5g/L
Arginine 100x (Sigma)	10g/L
Histidine 100x (Sigma)	10g/L
Uracil 20x (Sigma)	2g/L
Leucine 100x (Sigma)	10g/L

Additional supplements:

Fluoroorotic acid (FOA) (Melford Laboratories) was added to media at a concentration of 0.5g/500ml (1x) or 1g/500ml (2x).

Thiabendazole (TBZ) (Sigma) was added to media at concentrations of 10μg/ml, 15μg/ml or 20μg/ml in DMSO.

Nourseothricin (cloNAT) (Werner BioAgents) was added to media at a concentration of 0.4mg/ml.

Geneticin (G418) (Gibco) was added to media at a concentration of 0.1mg/ml.

2.1.2 Cell counting

Coulter counter

Cell number was estimated using a Beckman Z2 Particle Count and Size Analyzer. 100μl of cells were mixed with 10ml Isoton II (Beckman Coulter) solution and counted according to manufacturers instructions.

Haemocytometer

Cells were also counted using a haemocytometer. A haemocytometer is a special microscope slide which has a grid etched onto the glass. This grid is on a region of the slide which is 0.1mm lower than the rest of the slide. The grid

consists of 25 large squares which are subdivided into 16 smaller squares. Once the coverslip has been applied to the slide, 10µl of cell culture is pipetted underneath. This creates a known volume of $0.1mm^3$. The number of cells/ml can be calculated by multiplying the number of cells in the 25 large squares by 1×10^4.

2.1.3 Cell Culture

For physiological experiments *S. pombe* cells are harvested during the exponential growth phase which is between 2×10^6 and 1×10^7 cells/ml. To generate these cultures a "loopful" of freshly growing yeast was inoculated into 10mls liquid media and incubated generally overnight at 25°C (due to the ts nature of many of the mutants used) until cells reach log phase. This pre-culture was then diluted in an appropriate volume for a suitable amount of time to reach the desired concentration. Flask size is important for optimal growth and is determined by the volume of culture required: 25ml flask for up to 10ml culture, 100ml flask for up to 50ml culture, 250ml flask for up to 125ml culture and 500ml flask for up to 250ml culture.

2.1.4 Auxotrophy

The most commonly used auxotrophic markers in *S. pombe* are uracil, leucine, arginine, histidine and adenine. These amino acids are used at a concentration of 100µg/ml. To test auxotrophy, cells are grown as single colonies on non-selective media and then replica plated to minimal media lacking the appropriate supplement. The plates are then incubated for 2-3 days and examined for growth.

2.2. Yeast Molecular Genetics

2.2.1 Mating and random spore analysis

Crosses were carried out on ME medium in order to nitrogen starve the cells and induce sporulation. A similar amount of cells from two strains of opposite mating types (h^+/h^-) were mixed together and incubated for 2-3 days at 25°C. The cells were checked for the presence of ascii containing four spores by light microscopy. Cells were resuspended in 500µl of filter sterilised dH_2O containing 5µl of glusulase (NEN) and incubated overnight at 37°C. Glusulase digests the acsus wall and vegetative cells so that only the spores remain. The spores were plated on selective media at dilutions of 1:100 and 1:1000 and grown at 25°C or 32°C until colonies are formed.

2.2.2 *S. pombe* transformations

Electroporation

A 50ml culture of cells in log phase ($5x10^6$ to $1x10^7$ cells/ml) was harvested at 3000rpm for 3 minutes in a Sorvall Legend RT benchtop centrifuge. The pellet was washed once in 20ml ice-cold 1.2M sorbitol (Sigma) and then three times in 10ml 1.2M ice-cold sorbitol. The pellet was resuspended in 1.2M ice-cold sorbitol to a concentration of $1x10^9$ cells/ml. 200µl of cells were mixed with 100ng (plasmids) and/or 10µg (linear fragments) of DNA in an ice-cold cuvette. Cells were pulsed using a Bio-Rad Gene Pulser II at a setting of 2.25kV, 200Ω and 25µF. Immediately following pulsing, 1ml of 1.2M ice-cold sorbitol was added. Cells were spread at various dilutions onto selective media using sterile glass beads and incubated at 25°C or 32°C until colonies appeared.

Lithium acetate transformation

A 50ml culture of log phase cells was harvested as before. Cells were washed in 10ml 0.1M lithium acetate pH4.95 (Sigma), resuspended in 10ml 0.1M lithium acetate pH4.95 and incubated for between 30 minutes and 1 hour at 32°C. Cells were pelleted and resuspended at a concentration of $1x10^9$/ml in 0.1M lithium

acetate pH4.95. 1µg of DNA was added to 150µl of cells, mixed and then 370µl 50% PEG 3350 (Sigma) dissolved in TE was added. Cells were again incubated for between 30 minutes and 1 hour at 32°C, heatshocked for 20 minutes at 42°C then pelleted and resuspended in sterile dH_20. Cells were spread at various dilutions onto selective media using sterile glass beads and incubated at 25°C until colonies appeared. When transforming linear DNA fragments, after heatshocking cells are allowed to recover for a few hours to overnight in liquid non-selective media and then plated on selective plates.

2.2.3 Serial dilution assay

To assay the growth of different *S. pombe* strains on different media cells were taken from a plate and counted using a Beckman Z2 Particle Count and Size Analyzer. Serial dilutions of 1:10 were made in sterile microtitre plates in dH_2O starting with $5x10^6$ cells/ml and plated on the appropriate media using a multi-pin apparatus. Cells were then incubated at the desired temperature until colonies were formed.

2.2.4 Centromere silencing assay

Wild type cells which have the *ade6*$^+$ gene inserted into centromeric outer repeats are red when grown under restricted adenine conditions. This is due to transcriptional repression which causes the accumulation of amino-imidazole ribonucleotide (AIR) (Fisher, 1969). Mutants that alleviate silencing at the outer repeats are white due to alleviation of silencing. This assay can also be carried out using the insertion of the *ura4*$^+$ gene. In wild type cells, *ura4*$^+$ expression is repressed and cells are able to grow well on counter-selective media containing FOA. Mutant cells grow well on media lacking uracil but are unable to grow on media containing FOA (Boeke et al., 1984). Cells are spotted onto appropriate plates as described in 2.2.3.

2.2.5 *S. pombe* expression vectors

ars vectors

S. pombe vectors contain a bacterial origin of replication and selectable marker as well as a yeast selectable marker and an autonomous replication sequence (ars) or equivalent. Budding yeast markers are frequently used such as LEU2 which complements a mutation in the leu1 gene of *S. pombe.* Plasmids based on the *S. cerevisiae* origin 2μ are mitotically unstable, have a low copy number, are more prone to rearrangements and are more difficult to recover from fission yeast than plasmids containing *S. pombe* ars1. However, the copy number of ars1 vectors may vary as they can produce polymers with variable numbers of repeats.

Inducible vectors

The most commonly used inducible vectors in *S. pombe* are the pREP vectors. These contain the thiamine responsive promoter of the *nmt1*$^+$ gene which is repressed in the presence of thiamine and expressed in the absence of thiamine and gives full induction after around 16 hours (Maundrell, 1993). The *nmt1* promoter has been mutated to give different levels of expression as the fully induced level is very high. pREP1 and pREP3X have the highest levels of expression, pREP41X has a 15 fold lower level than pREP1 and pREP81X has an 80-fold reduction in expression level than pREP1 (Basi et al., 1993).

2.2.6 Plasmid recovery from *S. pombe*

To recover plasmids from fission yeast transformed with a genomic library, single colonies were grown at 32°C for 3 days in PMG media lacking the appropriate auxotrophic marker to select for colonies retaining the plasmid. Cells were harvested by centrifugation at 3000rpm for 2 minutes. The pellet was resuspended in SP1 containing 1mg/ml zymolyase 100-T (MP Biomedicals) and incubated for 1 hour at 37°C. Cells were harvested again and resuspended in Buffer P1 from the Qiagen Plasmid Miniprep kit. The Qiagen miniprep protocol was then followed as per manufacturers instructions and the DNA eluted in 30μl dH_2O. 10μl of the recovered DNA was transformed into 30μl DH5α competent

cells (Invitrogen) and plated on selective LB plates containing 30μg/ml ampicillin. As a low yield of plasmid DNA occurs from this initial rescue, a miniprep was performed on colonies from this first transformation and the DNA was then re-transformed into DH5α competent cells as before.
SP1: 1.2M sorbitol, 50mM sodium citrate, 50mM Na_2HPO_4.7H2O, 40mM EDTA, pH 5.6.

2.3 DNA protocols

2.3.1 Genomic DNA Isolation

A 5ml stationary phase culture was harvested at 3000rpm for 5 minutes. The pellet was resuspended in 250μl SP1 buffer containing 0.4mg/ml zymolyase 100-T (MP Biomedicals) and incubated for 30 to 60 minutes at 37°C. The cells were then pelleted at 13000rpm in a microfuge for 15 seconds and the pelleted resuspended in 0.5ml TE, 50μl 10% SDS and vortexed. 165μl 5M potassium acetate was then added and the samples incubated on ice for 30 minutes. After centrifugation at 13000rpm at 4°C for 10 minutes, the supernatant was added to 0.75ml iospropanol, incubated on dry ice for 5 minutes and then centrifuged as before. The pellet was resuspended in 0.3ml TE containing 10μg/ml of RNase A (Roche). After 30 minutes at 37°C the sample was then extracted with phenol/chloroform and precipitated by addition of 2-3 volumes of ethanol and 1/10 volume of 3M sodium acetate. The pellet was resuspended in 20μl TE and stored at -20°C.

SP1: 1.2M sorbitol, 50mM sodium citrate, 50mM sodium phosphate, 40mM EDTA, pH 5.6

2.3.2 Genomic DNA Isolation for Whole-Genome Sequencing

2ml stationary phase cultures for 10 individual isolates of the mutant strain were combined in equal cell numbers and harvested at 3500rpm for 2 minutes. The

pellet was resuspended in 1ml SP1 buffer containing 1µl of β-mercaptoethanol and incubated for 20 minutes at 25°C. The cells were spun again at 3500rpm for 2 minutes and resuspended in 1ml SP1 buffer containing 0.1mg/ml zymolyase 100-T (MP Biomedicals) and incubated for 1 to 2 hours at 37°C. The cells were then pelleted at 13000rpm in a microfuge for 1 minute and the pellet resuspended in 0.3ml 4M Guanidine Hydrochloride. The cells were incubated at 65°C for 10 minutes and allowed to cool to room temperature before 0.3ml cold ethanol was added. The cells were consequently centrifuged at 10000rpm, at 4°C for 5 minutes. The pellet was resuspended in 0.6ml TE containing 10µg/ml of RNase A (Roche). After 1 hour at 37°C Proteinase K (Roche) was added to a concentration of 10µg/ml and incubated at 65°C for a further 1 hour. The sample was then extracted with phenol/chloroform and precipitated by addition of 6/10 volume of isopropanol and 1/10 volume of 3M sodium acetate with incubation at room temperature for 30 minutes. The pellet was resuspended in 50µl TE and stored at -20°C.

SP1: 1.2M sorbitol, 50mM sodium citrate, 50mM sodium phosphate, 40mM EDTA, pH 5.6

2.3.3 Rapid DNA isolation for PCR

A single colony of *S. pombe* was suspended in 30µl SPZ buffer and incubated at 37°C for 1 hour. Crude DNA was used in PCR analysis undiluted.
SPZ: 1.2M sorbital, 100mM sodium phosphate, 2.5mg/ml zymolyase-100T (MP Biomedical)

2.3.4 Agarose gel electrophoresis

Agarose gel electrophoresis was used to analyse the size of DNA fragments. Agarose (Melford) at a final concentration of 1% was dissolved in 1 x TBE buffer by heating in a microwave. Once cooled, ethidium bromide (Sigma) was added

to a concentration of 0.03µg/ml. DNA samples were loaded in Orange G loading buffer and visualised under a UV transilluminator. Data capture was achieved using the U:Genius system (Syngene) .

10xTBE: 108g Trizma base, 55g boric acid, 9.3g EDTA

Loading buffer: 15% ficoll in TE, Orange G

2.3.5 Polymerase Chain Reaction (PCR)

PCR reactions were carried out as follows in 0.2µl thin walled PCR tubes, initially using Roche Taq which was later replaced with FastStart Taq: template DNA, 10pM primer, 2.5mM dNTPs , 1x PCR buffer, 0.5U Taq/FastStart Taq (Roche), dH2O. When precise amplification was required, in the case of cloning and fusion PCR, Platinum *Pfx* Taq polymerase from Invitrogen was used as per manufacturer's instructions. All reactions were carried out using either a PTC-225 thermal cycler (MJ Research) or a T3000 Thermocycler (biometra). The following programs were used as indicated.

Ura program: 94°C for 4 minutes, (94°C for 30 seconds, 55°C for 30 seconds, 72°C for 1 minute), 29 cycles, 72°C for 5 minute.

PFX-1min: 94°C for 5 minutes, (94°C for 30 seconds, 55°C for 30 seconds, 68°C for 1 minute), 29 cycles, 68°C for 5 minutes.

Fast-6A: 95°C for 6 minutes, (95°C for 30 seconds, 55°C for 30 seconds, 72°C for 1 minute), 29 cycles, 72°C for 7 minutes.

2.3.6 Fusion PCR

Fusion PCR constructs were produced in a two-step reaction in 0.2ml thin walled PCR tubes (Yu 2004). The first step involved complementary regions of homology on the fragments being used as primers for DNA polymerase to produce full length constructs. The second step is a PCR reaction using nested primers to amplify the fusion construct.

First Step: Three DNA fragments with homology-containing overhangs added in 1:3:1 molar ratio, 0.2mM dNTPs , 2x PCR buffer, 1mM $MgSO_4$, and 0.625U Platinum *Pfx* Taq (Invitrogen), dH2O up to 25µl.

The reaction was run on:

PFX-Fus: 94°C for 2 minutes, (94°C for 30 seconds, 58°C for 8 minutes, 68°C for 5 minutes), 14 cycles, 68°C for 5 minutes.

Second Step: 0.2µl 1st reaction template, 10pM primer, 0.3mM dNTPs , 2x PCR buffer, 1mM $MgSO_4$, and 1.25U Platinum *Pfx* Taq (Invitrogen), dH2O up to 50µl. The reaction was run on:

PFX-Long: 94°C for 5 minutes, (94°C for 30 seconds, 55°C for 30 seconds, 68°C for 5 minutes), 34 cycles, 68°C for 5 minutes.

2.3.7 Quantitative (q)PCR

PCR reactions were carried out as follows in 0.2µl bright white 96 well PCR plates: 5µl of 1:10 dilted template DNA, 0.4µl of 10mM primer, 0.4µl of 1mM Fluorephen , 10µl qPCR mix (Sigma), up to 20µl with dH2O. All reactions were carried out using the Lightcycler 480 (Roche). The following program was used:

Normal qPCR Template

Stage	Temperature (°C)	Time (mins:secs)	Ramp Speed (°C/s)
Hotstart	94	02:00	4.4
Quantification (50 Cycles)	95	00:15	4.4
	55	00:30	2.2
	72	00:30	4.4
Melting Curve	95	00:30	4.4
	55	00:30	2.2
	95	Until temp reached	0.11
Cooling	40	01:00	4.4

2.3.8 Mutagenic PCR

Reactions were carried out using the Genemorph II Random Mutagenesis Kit (Stratagene) in 0.2ml thin walled tubes: 5µl 10x Mutazyme II reaction buffer (Stratagene), 2.5U of Mutazyme II DNA Polymerase (Stratagene), 1µl of each primer (10µM), 500-1000ng template DNA and dH_2O up to 50µl. Reactions were run on the following program in a PTC-225 thermal cycler (MJ Research):
RanMut30: 95°C for 2 minutes, (95°C for 1 minute, 55°C for 1 minute, 72°C for 1 minute), 29 cycles, 72°C for 10 minutes.

2.3.9 Site-directed Mutagenesis

Reactions were carried out using the QuikChange II Site-Directed Mutagenesis Kit (Stratagene) in 0.2ml thin walled tubes: 5µl reaction buffer (Stratagene), 2.5U of Pfu Ultra HF DNA Polymerase (Stratagene), 1µl dNTP mix (Stratagene), 125ng of each primer, 10ng plasmid DNA and dH_2O up to 50µl. Reactions were run on the following program in a PTC-225 thermal cycler (MJ Research):
MUT-4.5K: 95°C for 1 minute, (95°C for 50 seconds, 60°C for 50 seconds, 68°C for 4 minutes 30 seconds), 17 cycles, 68°C for 7 minutes.

2.3.10 Sanger Sequencing

Reactions were set up as follows: 2µl ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kit v 3.0 (Applied Biosystems), 3.2pmol/µl primer, template DNA as recommend by manufacturers (1-1000ng for PCR products, 200-500ng for dsDNA) and dH_2O up to 20µl. Reactions were run on the following program in 0.2µl thin walled PCR tubes in a PTC-225 thermal cycler (MJ Research): 95°C for 5 minutes, (95°C for 30 seconds, ramp 1°C per second to 55°C, 55°C for 15 seconds, ramp 1°C per second to 64°C), 25 cycles, 64°C for 4 minutes. Samples were then sent to the central sequencing service for analysis.

2.3.11 Illumina Whole-Genome Sequencing

Genomic DNA samples were generated as in 2.3.2. Samples were sent to the GenePool sequencing service where the DNA was processed and run on the Illumina sequencing platform GAII for single-end reads or GAIIx for paired-end reads as specified. Bioinformatic processing was also done by GenePool where the sequences were aligned on the *S. pombe* reference sequence from the Sanger Institute using the Mapping and Assembly with Quality (Maq) program and a list of SNPs was produced.

2.4 RNA protocols

2.4.1 Total RNA isolation

A 50ml culture of cells in log phase was harvested then resuspended in 1ml TE and transferred to an eppendorf tube. Cells were pelleted and resuspended in 600µl extraction buffer, 600µl phenol:chloroform 5:1 (Sigma) and 600µl 425-600 micron acid-washed beads (Sigma). Cells were lysed on a multi-head vortexer at maximum speed for 30 minutes at 4°C. The samples were spun at 13,000rpm at 4°C for 5 minutes and the supernatant transferred to a fresh tube. Samples were extracted with phenol:chloroform and subsequently with chloroform then precipitated by adding 3 volumes of ice-cold 100% ethanol and centrifuging at 13,000rpm for 15 minutes. The pellet was resuspended in 30-50µl dH2O.

Alternatively, RNA was made using a Qiagen RNeasy Miniprep or Midiprep kit as per manufacturer's instructions. RNA was quantified using a Nanodrop spectrophotometer.

Extraction buffer: 50mM Tris-HCl pH7.5, 10mM EDTA pH8, 100mM NaCl,1% SDS

2.4.2 Small RNA isolation

A 50ml culture of cells in log phase was harvested then resuspended in 1ml TE and transferred to an eppendorf tube. Cells were pelleted and resuspended in

600µl extraction buffer, 600µl phenol:chloroform 5:1 (Sigma) and 600µl 425-600 micron acid-washed beads (Sigma). Cells were lysed on a multi-head vortexer at maximum speed for 30 minutes at 4°C. The samples were spun at 13,000rpm at 4°C for 5 minutes and the supernatant transferred to a fresh tube. Samples were extracted with phenol:chloroform 5:1, followed by 25:24:1 and subsequently with chloroform. Large rRNA, mRNA and genomic DNA were removed by precipitation with 10% polyethylene glycol 8000 and 0.5M sodium chloride. The samples were incubated on ice for 30 minutes then spun at 13,000rpm for 20 minutes. The pellet was dissolved in 60µl dH2O. The supernatant containing the siRNAs was then precipitated by the addition of 3 volumes of 100% ethanol and 1/10 volume sodium acetate and incubation at -20°C overnight. The samples were spun at 13,000rpm for 20 minutes and the pellets were washed in 95% ethanol. The samples were again spun for 10 minutes at 13,000rpm and resuspended in 35µl dH2O. The samples were stored at -80°C.
Extraction buffer: 50mM Tris-HCl pH7.5, 10mM EDTA pH8, 100mM NaCl,1% SDS

2.4.3 Small RNA isolation from immunoprecipitated Ago1 protein

A 2L 4xYES culture of cells in log phase at a concentration of 2 x 10^8 cells/ml was spun at 10000rpm for 15 minutes at 4°C. The cells were washed 3 times in 10ml PBS before being taken dry and drop-frozen in liquid nitrogen. Cells were lysed by grinding in liquid nitrogen using a Retsch mortar grinder for 30 minutes. Powder was collected in 10g aliquots. 10ml of cold lysis buffer was added to 10g of powder and the sample was solubilised by rotation for 30 minutes at 4°C. The solubilised extract was then centrifuged at 4000rpm for 5 minutes at 4°C. The supernatant was spun again at 15000rpm for 10 minutes at 4°C. The supernatant was filtered through a 1.6µm GD/X Whatmann syringe filter. During solubilisation 4µl of protein G dynabeads (Invitrogen) were coupled to 8µl M2 monoclonal anti-flag antibody (Sigma F1804) through rotation for 1 hour at 4°C before being resuspended in 10µl lysis buffer. The 10µl of coupled dynabeads was added to the supernatant and rotated for 60 minutes at 4°C. Beads were collected

using a magnet, the supernatant was discarded and the beads were washed in 1ml lysis buffer. The beads were transferred to an eppendorf tube and washed with 0.5ml of lysis buffer containing no protease inhibitors. The supernatant was again discarded and the beads were resuspended in 300µl 50:50 TENS:water containing 40µg Proteinase K (Roche). The sample was shaken at 37°C for 2 hours. Samples were extracted with phenol:chloroform 25:24:1 and subsequently with chloroform. The supernatant containing the siRNAs was then precipitated by the addition of 3 volumes of 100% ethanol and 1/10 volume sodium acetate and incubation at -20°C overnight. The samples were spun at 13,000rpm for 20 minutes and the pellets were washed in 95% ethanol. The samples were again spun for 10 minutes at 13,000rpm and resuspended in 15µl dH2O. The samples were stored at -80°C.

Lysis buffer: 50mM HEPES-NaOH pH7.5, 150mM NaCl, 1mM MgCl2, 0.1% NP-40 (Sigma), 5mM DTT (Invitrogen), 1x Complete EDTA-free; Protease Inhibitor Cocktail Tablet (Roche), 0.5mM PMSF, 1/100 SUPERase-In (Ambion)

2.4.4 Denaturing Polyacrylamide Gel Electrophoresis

15µl of the siRNA samples was diluted in 10µl FDE sample buffer, denatured for 3 minutes at 95°C then stored on ice prior to loading. siRNA samples were run out on a 12% 16 x 18cm polyacrylamide gel made using the Sequagel system (National Diagnostics). Gels were run in the Hoefer SE600 Ruby apparatus at 300V first pre-run for 30 minutes and then for ~2 hours and 30 minutes once the samples were added (until the bromophenol blue dye front was approximately 2 cm from the bottom). The gel was rinsed in 20 x SSC then transferred to a membrane by northern blotting.

FDE Sample Buffer: 0.5M EDTA pH8, 10mg xylene cyanol,10mg bromophenol blue, made up to 10ml with deionised formamide

2.4.5 Northern Blotting

Gels were blotted by capillary transfer for at least 16 hours in 20 x SSC using Hybond-NX (Amersham). The membrane was crosslinked twice at 1200 joules in a UV crosslinker (Stratagene).

Hybridisation and probe preparation

Membranes were pre-hybridised for a minimum of 1 hour at 42°C in 20ml Church buffer. During this time DNA probes were labelled using the T4 Polynucleotide Kinase (PNK) (NEB) 5' end labelling kit:

siRNA Probe: 4pmol oligo (IK8, IK9, IK10), 1x T4 PNK buffer, 10U T4 PNK, 8pmol [γ-^{32}P] ATP and up to 10µl with dH2O

snoRNA58 Probe: 100pmol oligo (IK5, IK7), 1x T4 PNK buffer, 10U T4 PNK, 1.6pmol [γ-^{32}P] ATP and up to 10µl with dH2O

The reaction was incubated for 1 hour at 37°C and unincorporated radionucleotide was removed using an illustra ProbeQuant G-50 Micro Column (GE Healthcare). The probes were added directly to the Church buffer. Membranes were hybridised overnight at 42°C and subsequently washed three times for 20 minutes at 42°C in 2xSSC/0.2%SDS. The membranes were sealed in a bag and placed in a cassette with a phosphoscreen overnight. The signals were visualised on a Storm phosphoimager with ImageQuant TL v 2005 (Amersham).

Church buffer: 0.5M sodium phosphate pH 7.2, 1mM EDTA, 7% SDS

2.4.6 Reverse Transcriptase (RT)-PCR

RNA samples were obtained as described in 2.4.1. 1µg of total RNA was aliquoted into a 0.2ml thin walled PCR tube. 1µl (2U) Turbo DNAse (Ambion) and 1µl of 10x Turbo DNAse buffer were added and samples made up to a final volume of 10µl with dH_2O. Samples were incubated for 1 hour at 37°C. 2µl of 10mM dNTP (Roche) and 2µl of 100mM random hexamer were added and

made up to 26µl with dH2O. Samples were incubated at 65°C for 10 minutes and then placed on ice. Next 8µl of 5x First Strand buffer (Invitrogen), 2µl of 0.1M DTT (Invitrogen) were added and made up to 38µl with dH2O. Samples were then mixed and split into two aliquots. 1µl (200U) Superscript III Reverse Transcriptase (Invitrogen) was added to one set of samples whilst the 1µl of dH2O was added to the second set. Samples were then incubated at 50°C for 60 minutes and 70 °C for 15 minutes and then placed on ice. 1µl of this cDNA was used as a template in a 20µl PCR reaction.

2.5 Protein Protocols

2.5.1 *S. pombe* protein extraction and immunoprecipitation

A 200ml 4xYES culture of cells in log phase at a concentration of 2 x 10^8 cells/ml was spun at 10000rpm for 15 minutes at 4°C. The cells were washed twice in 10ml cold PBS, resuspended in $1/5^{th}$ volume of PBS and drop-frozen in liquid nitrogen. Cells were lysed by grinding in liquid nitrogen using a manual pestle and mortar for 10 minutes. 5ml of cold lysis buffer was added and the sample was solubilised by rotation for 30 minutes at 4°C. The solubilised extract was then centrifuged at 4000rpm for 10 minutes at 4°C. The supernatant was spun again at 13000rpm for 30 minutes at 4°C, 50µl was taken as an input sample and 50µl sample buffer was added to the input. During solubilisation 8µl of protein G dynabeads (Invitrogen) were coupled to 8µl M2 monoclonal anti-flag antibody (Sigma F1804) through rotation for 1 hour at 4°C before being resuspended in 10µl lysis buffer. The 10µl of coupled dynabeads was added to the remaining supernatant and rotated for 60 minutes at 4°C. The beads were collected using a magnet, the supernatant was discarded and the beads were washed 4 times in 1ml lysis buffer. The beads were resupended in 40µl sample buffer. Samples were boiled for 5 minutes at 95°C then spun in a microfuge for 30 seconds to remove beads and cell debris. Samples were then either immediately loaded on an SDS-PAGE gel or stored at -20°C until required.

Lysis buffer: 50mM HEPES pH7.5, 150mM NaCl, 5mM EDTA, 0.1% NP-40 (Sigma), 5mM DTT (Invitrogen), 1x Complete EDTA-free; Protease Inhibitor Cocktail Tablet (Roche), 0.5mM PMSF (Sigma), 1x Protease Inhibitor Cocktail (Sigma)
2x Sample buffer: 2% SDS, 50mM Tris-HCl pH6.8, 2mM EDTA, 10% glycerol, 0.03% bromophenol blue, 2% β-mercaptoethanol

2.5.2 SDS-PAGE

Proteins were separated on 1mm thick SDS polyacrylamide gels using the Hoefer minigel apparatus. Resolving gel was poured at concentrations of either 6% or 8% depending on the size of the protein. Stacking gel (5%) was poured on top and gels were run at 200V for around 40 minutes in 1 x running buffer. After running the gels were transferred to nylon membrane for Western blotting.

Polyacrylamide gel 10ml: 2.5ml 1.5M Tris-HCl pH8.8, 100µl 10% SDS, 100µl 10% ammonium persulphate, 10µl TEMED, 30% acrylamide/bis mix (Sigma) either 2ml or 2.7ml depending on concentration, up to 10ml with dH_2O.

Stacking gel 100ml: 17ml 30% acrylamide/bis, 12.5ml 1M Tris-HCl pH6.8, 1ml 10% SDS, 69.5ml dH_2O.

5xRunning Buffer: 30g Tris Base, 144g glycine, 5g SDS in 1L dH2O.

2.5.3 Western analysis

Proteins were transferred onto Protran nitrocellulose membrane (Schleicher and Schuell) using a Hoefer semi-dry electroblotter. Before use, the membrane was floated on top of dH_2O. The blotting apparatus was assembled with 6 pieces of 3MM paper soaked in blotting buffer, the membrane stacked on top, then the gel, then another 6 pieces of 3MM paper. Any bubbles were removed by rolling over the top of the stack with a glass test-tube. Transfer was carried out for 1-2hours at 65mA per gel. After transfer, the membrane was washed briefly in dH_2O and stained with Ponceau S (Sigma) solution to confirm transfer. The

membrane was washed in PBS and placed in blocking buffer for 1 hour at room temperature. Primary antibody was then added at appropriate concentration in PBS + 0.1% Tween 20 and incubated for 1 hour at room temperature. Membranes were then washed twice in PBS + 0.1% Tween 20 for 15 minutes each and then the secondary HRP-conjugated antibody was added in blocking buffer. The secondary antibody was also incubated for 1 hour at room temperature. The membrane was washed again as before and then rinsed briefly in PBS. Proteins were detected using an Enhanced Chemi-Luminescence kit (Amersham) as per manufacturer's instructions. The blot was exposed to Bio-Max Light film (Kodak) for between 30 seconds and 10 minutes.
Blotting buffer: 20ml 5 x SDS running buffer, 20ml methanol, 20ml dH_2O
Blocking buffer: 5% Marvel dried nonfat milk, PBS+0.1% Tween 20.

2.5.4 Chromatin Immunoprecipitation (ChIP)

A 50ml culture was grown in YES to log phase between $5x10^6$ cells/ml and $1x10^7$ cells/ml. Cells were transferred to a falcon tube and fixed with 1% formaldehyde in the fume hood for 15 minutes. The samples were centrifuged for 2 minutes at 4°C in a Sorvall Legend RT benchtop centrifuge. The cells were then washed twice in 20ml ice-cold PBS and transferred to a round-bottomed 2ml screw-capped tube in 1ml PBS. Cells were spun at 3500rpm for 2 minutes, the supernatant discarded and the pellet frozen on dry ice for storage at -80°C. The cells were thawed on ice and added 350µl lysis buffer containing 1mM PMSF (Sigma) and 1/100 Protease Inhibitor Cocktail (Sigma) and 500µl 425-600 micron acid-washed beads (Sigma). Cells were lysed using a mini beadbeater (Biospec Products) on maximum rotation speed for 2x 2 minutes with a 1 minute incubation on ice between rotations. A small hole was made in the base of the tube using a heated needle and the tube was placed inside an eppendorf tube and spun for 1 minute at 1000rpm in order to transfer the supernatant. The samples were then sonicated using a BioRuptor Sonicator (Diagenode) for a total of 15 minutes. Samples were then pelleted in a microfuge at 13,500rpm for 5mins at

4°C and the supernatant transferred to a fresh tube. Samples were spun again for 15mins at 4°C and the supernatant transferred to a fresh tube. While samples are centrifuging, Protein G agarose beads (Roche) was washed with lysis buffer 3 times and finally resuspended as a 50:50 beads:lysis buffer slurry. The crude lysates were pre-cleared using 25ul Protein G for 1 hour at 4°C. Beads were pelleted at 2,000rpm for 2 minutes in a microfuge. The lysate was transferred to a fresh tube with 30ul retained as the crude control and frozen at -20°C. 1µl of H3K9me2 monoclonal antibody 5.1.1. and 25ul Protein G beads were added to the cell lysate and incubated at 4°C overnight. Following the antibody incubation, beads were centrifuged at 2000rpm for 2 minutes in a microfuge. Beads were then washed in 1ml each of the following ice-cold buffers for 5 minutes at 4°C: lysis buffer, lysis buffer + 0.5 M NaCl, wash buffer, TE pH8. 250ul of TES was added to the beads and 220ul TES added to the crude extract. All samples were incubated at 65°C overnight to reverse the crosslinks. Samples were centrifuged for 2 minutes at 2000rpm and the supernatant removed to a fresh tube, discarding the beads. 450ul TE and 30ul 10mg/ml Proteinase K (Roche) were added to each sample and incubated for 2 hours at 37°C. Samples were then extracted with phenol:chloroform 25:24:1 followed by chloroform. The DNA was precipitated with 3M sodium acetate pH5.5 and 100% ethanol. The samples were incubated for 30 minutes on dry ice and then centrifuged at 13500rpm for 30 minutes at 4°C. The supernatant was removed and pellets were dried under a hood then ChIP samples were resuspended in 30µl TE and crude samples in 300µl TE. Samples were then analysed by basic PCR or qPCR. For all basic PCR reactions the 'Ura' program was used and the qPCR program can be found in 2.3.7.

Lysis buffer: 50mM HEPES-KOH pH7.5, 140mM NaCl, 1mM EDTA, 1% Triton-X100, 0.1% sodium deoxycholate

Wash buffer: 10mM Tris-HCl pH8, 0.25M LiCl, 0.5% NP-40 (Sigma), 1mM EDTA, 0.5% sodium deoxycholate

TE: 10mM Tris-HCl pH8, 1mM EDTA

TES: 50mM Tris-HCl pH8, 10mM EDTA, 1%SDS

2.6 Bacterial Protocols

2.6.1 Media

LB medium per litre:	Bacto tryptone	10g
	Bacto yeast extract	5g
	Sodium chloride	10g
LB agar per litre:	Bacto tryptone	10g
	Bacto yeast extract	5g
	Sodium chloride	10g
	Bacto agar	15g

Antibiotics:

Kanamycin 50mg/ml

Ampicillin 100mg/ml

Carbenicillin 50mg/ml

Chloramphenicol 20mg/ml

2.6.2 Transformation

DH5α (Invitrogen) cells were transformed as per manufacturer's instructions with between 1 and 5μg plasmid DNA. In brief, between 30-100μl of cells were mixed with DNA, incubated on ice for 30 minutes, heatshocked at 42°C for 45 seconds and returned to ice for 1 minute. Cells were then grown at 37°C for 1 hour with the addition of between 100-400μl of SOC medium. Cells were plated on media or grown in liquid supplemented with the appropriate antibiotic.

SOC 1L: Bacto tryptone 20g, bacto yeast extract 5g, sodium chloride 20g, 10ml 250mM potassium chloride, 5ml 2M magnesium chloride, 20ml 1M glucose pH 7

2.6.3 Plasmid construction

PCRs for cloning were carried out using Platinum *Pfx* Taq (Invitrogen). Restriction enzymes were obtained from New England Biolabs. DNA fragments were recovered and purified using a Gel Extraction kit (Qiagen). Ligations were performed using T4 DNA ligase (Promega) and incubated at 4°C overnight with insert:vector ratios of 1:1 and 3:1. Ligations were transformed into DH5α (Invitrogen) cells and plated on media containing appropriate antibiotic.

2.6.4 Plasmid miniprep

Single bacterial colonies were grown in 5ml LB plus appropriate supplement overnight at 37°C. Cells were harvested and plasmid prep performed using a QIAprep miniprep kit (Qiagen) according to manufacturer's instructions.

2.7 Antibodies

2.7.1 ChIP

mouse anti-diMeH3K9 1:300 (m5.1.1., gifted from Takeshi Urano lab)

2.7.2 CoIP Western Analysis

M2 FLAG-HRP 1:1000 (Sigma)

12CA5 HA 1:1000 (abcam)

2.8 Primers

Name	Sequence 5'-3'	Description
Ade_Int_R	CTTCAATGGTGTAGTGACCTG	*ade6*$^+$ Internal primers for testing partial deletion
Ade_Int_F	TGCAACTCTGCGATGCATTC	*ade6*$^+$ Internal primers for testing partial deletion
rdp_3900_F	GACCTGAGATCAATTATGGCT	Testing rdp1 flag- tag
Flag_R	CACCGTCATGGTCTTTGTAG	Testing rdp1 flag- tag
clr4$^+$80_R	GTTTGACAGCTCCATTACGAT	Testing *clr4*$^+$ flag- tag
clr4-200_F	CTCTGAAATTGAACACATCGA	Testing *clr4*$^+$ flag- tag
3-Ade6(OTR)	GGCCACCATAGACATAACTG	Testing for *ade6*$^+$ at otr
5-OTR-ade6	CTACTCTTCTCGATGATCCTGTA	Testing for *ade6*$^+$ at otr
Ura 3'UTR_F	CATGCTCCTACAACATTACCACAATC	Testing for *ura4* deletion
Ura 5'UTR_R	GCAAGGGCATTAAGGCTTATTTACAG	Testing for *ura4* deletion
MT1	AGAAGAGAGAGTTGAAG	Testing mating-type
MM	TACGTTCAGTAGACGTAGTG	Testing mating-type
MP	ACGGTAATCGGTCTTCC	Testing mating-type
dcr -150 F	GTATTCTGCTCGTGTGATTGTTC	Testing *dcr1* deletion
dcr in 150 R	CAACTTCACAGCAAGTAATGTC	Testing *dcr1* deletion
Ptef 100 R	CGTCAAGACTGTCAAGGAG	Testing *dcr1* deletion
ago1 3' F	GACCAATTCCAAACACTGTG	Testing ago1 deletion
ago1 3' R	CTTATTGCATGCAATCCATCAAAC	Testing ago1 deletion
Ura F1	CTGGGACAGCAATATCGTACTCCTGAA	Testing ago1 deletion
Ago1_nmt1_Long_F	GGTTTGGTATATATAAGCTTCCAACCGCCAAAGCGAATTGTCTTCAGCCAACTCGTCCTTTATGATTCAGAGTGAGTAGGGAATTCGAGCTCGTTTAAAC	Insertion of nmt1 promoter at Ago1
Ago1_nmt1_Long_R	AAGTTAGCTTTCAAAGTAATCTGCTTGCCCAACCCTCCATAACCGGGACGTAAAGC	Insertion of nmt1 promoter at Ago1

	TATTTCTGAGCTTGGTTTATACGACATGATTTAACAAAGCGACTATA	
Ago1_nmt1_int_F	CCGTCGATTTGGAACCTTTA	Testing nmt1 integration at Ago1
Ago1_nmt1_int_R	GACAAGAAAGGCCACGTGTT	Testing nmt1 integration at Ago1
Kan_Int_R	TGGTCGCTATACTGCTGTCG	Testing KanMX integration from pFA6a
pFA6a_R	CTCAAGAATAAGAATTTTCG	Testing KanMX integration from pFA6a
Rpb1_Gene_F	ATGGGTTTTGGAAACAGACG	Rpb1-1 Mutagenesis Fusion PCR
Rpb1_Gene_Nest	TACCCGTCATGGCATTAACA	Rpb1-1 Mutagenesis Fusion PCR
Rpb1_Gene_Fusion	TCAGCACTGAGCAGCGTAATCCAAAAGCATTGTCATCTTCA	Rpb1-1 Mutagenesis Fusion PCR
Rpb1_Plasmid_R	AGCTAAACGGCAACCACAAC	Rpb1-1 Mutagenesis Fusion PCR
Rpb1_ADH_Term_F	TGAAGATGACAATGCTTTTGGATTACGCTGCTCAGTGCTGA	Rpb1-1 Mutagenesis Fusion PCR
Rpb1_Kan_Term_R	AGTGCTGCGCATAAACCTTTGAATTCGAGCTCGTTTAAAC	Rpb1-1 Mutagenesis Fusion PCR
Rpb1_Down_Fusion	GTTTAAACGAGCTCGAATTCAAAGGTTTATGCGCAGCACT	Rpb1-1 Mutagenesis Fusion PCR
Rpb1_Down_R	TTGGGGTTTTCATCAAGAGG	Rpb1-1 Mutagenesis Fusion PCR
Rpb1_Down_Nest	GATTGGTAGCACAACGCAAA	Rpb1-1 Mutagenesis Fusion PCR
Rpb1-1_Seq_R	GGATGGACTTGTCGCTGAGT	Sequencing rpb1 G1443A
Rpb1-1_Mutagenic_F	GGACAATTAGCCCCAATGGCAACTGGCGCATTTG	Forward Mutagenic Primer for rpb1-G1443A
Rpb1-1_Mutagenic_R	CAAATGCGCCAGTTGCCATTGGGGCTAATTGTCC	Reverse Mutagenic Primer for rpb1-G1443A
Rpb1_Seq_F	TACCCGTCATGGCATTAACA	Sequencing rpb1 G1443A
Rpb1_SPZ_R	GGTTTGGCGATTATGGATTG	Sequencing rpb1 G1443A
Rpb1-1_ii_seq_R	TTGGGAAGTACCCATTCCTG	Sequencing rpb1 G1346A
Rpb1_Gene_Nest_ii	GCATAAATCTTACTGAAGCAATG	Rpb1-1 Mutagenesis Fusion PCR
Rpb1-1_ii_Mutagenic_F	CGTGGAAATTTTGCAAATTCTTGCTATTGAAGCTACGAGATCTGC	Forward Mutagenic Primer for rpb1-G1346A
Rpb1-1_ii_Mutagenic_R	GCAGATCTCGTAGCTTCAATAGCAAGAATTTGCAAAAT	Reverse Mutagenic Primer for rpb1-G1346A

	TTCCACG	
Rpb2_Long_F	ATTGTTGTATTTCACCTGTATTGTTGATTATTATCAGTAACCAAGGTCGTTAATTATGTAGAGAAGGTAATTTTTATTATATTACGCTGCTCAGTGCTGA	Insertion of KanMX on 3' of rpb2
Rpb2_Long_R	CGTAGTTCAATGAGATTTAGTAAAGAAATCATATCTAATAGCACATAAACGAATTTCAATTACTGATTGGGACAAATGGAGAATTCGAGCTCGTTTAAAC	Insertion of KanMX on 3' of rpb2
Rpb2_Kan_int_F	CCTACTCCCAAATCCAAGCA	Testing integration of KanMX at rpb2
Rpb2_Kan_int_R	ATATGTTGGCGGTGAGGAAG	Testing integration of KanMX at rpb2
Rpb3_Downstream_F	AAGGAGACGCTTTCGGTTCT	Rpb3 Fusion PCR with KanMX
Rpb3_Promoter_R	TCTGTCGGGATGAACAGAGA	Rpb3 Fusion PCR with KanMX
Fusion_Rpb3_Downstream_R	GTTTAAACGAGCTCGAATTCTGCGTCCGTTCATGTCATAC	Rpb3 Fusion PCR with KanMX
Fusion_Kan_Term_F	GTATGACATGAACGGACGCAGAATTCGAGCTCGTTTAAAC	Rpb3 Fusion PCR with KanMX
Fusion_ADH1_Term_R	CACGTGGTAGACTATGCTTCAAATTACGCTGCTCAGTGCTGA	Rpb3 Fusion PCR with KanMX
Fusion_ Rpb3_Gene_F	TCAGCACTGAGCAGCGTAATTTGAAGCATAGTCTACCACGTG	Rpb3 Fusion PCR with KanMX
Rpb3_Downstream_Nest_F	CGAAAGTGTTCTCTTTCACGA	Rpb3 Fusion PCR with KanMX
Rpb3_Promoter_Nest_R	AGCCTAATGCCGTAGATTCG	Rpb3 Fusion PCR with KanMX
Rpb3_Gene_F	TTGAAGCATAGTCTACCACGTG	Rpb3 Fusion PCR with KanMX
Rpb3_Gene_R	GGATTCAGAAACGCATATTACGA	Rpb3 Fusion PCR with KanMX
Rpb3_Centre_F	TTTTGGATCGGCGAGAATAG	Sequencing rpb3
Rpb3_Centre_R	CTATTCTCGCCGATCCAAAA	Sequencing rpb3
Rpb3_Down Long	ACCGAAAATTAAAAAGATATCAAAACGACAATTCCGAAAGTGTTCTCTTTCACGAAATTATCTAAAGCTATAATAAATCCGAATTCGAGCTCGTTTAAAC	Insertion of NAT on C-term of Rpb3

Rpb3_Term_Long	ACTATGCTTCAATTCTTTTC TAAAAATTATCCATTTTCA TCCTTTATATGTTTGTTGTT ATAAAAAATACTCTAGATG CGATTACGCTGCTCAGTGC TGA	Insertion of NAT on C-term of Rpb3
Rpb11_Term_Long	TTTCAAAAAATACAGAATA CCCTCTGTATGCATCTTTAT TACAAAAATTTTGACTTTC GTTTTCGCAACTTCTGTAA TTTATTACGCTGCTCAGTG CTGA	Insertion of NAT on C-term of Rpb11
Rpb11_Down_Long	GAAAAAAAAAGCAATATC AGAGAATAATCCAGGTAC CTAGAAAAACTTTCATCGC CATTGCTGCAGTTTTCTAG TTGGTTGAATTCGAGCTCG TTTAAAC	Insertion of NAT on C-term of Rpb11
Rpb11_Downstream_R	AGAATCGATGCTCTGCCCT A	Rpb11 Fusion PCR with KanMX
Rpb11_Fusion_Downstream_F	GTTTAAACGAGCTCGAATT CTTTCGTTTTCGCAACTTCT G	Rpb11 Fusion PCR with KanMX
Rpb11_Promoter_F	TGAGCCGATTTCCAACAGA T	Rpb11 Fusion PCR with KanMX
Rpb11_Fusion_Gene_R	TCAGCACTGAGCAGCGTA ATAAAAGCAGCCGTGGAT TTTA	Rpb11 Fusion PCR with KanMX
Rpb11_Fusion_Kan_Term_R	CAGAAGTTGCGAAAACGA AAGAATTCGAGCTCGTTTA AAC	Rpb11 Fusion PCR with KanMX
Rpb11_Fusion_ADH1_Term_F	TAAAATCCACGGCTGCTTT TATTACGCTGCTCAGTGCT GA	Rpb11 Fusion PCR with KanMX
Rpb11_Downstream_Nest_R	AAACTTTCATCGCCATTGC T	Rpb11 Fusion PCR with KanMX
Rpb11_Promoter_Nest_F	TAAGTGTCCGTTCACGTTC G	Rpb11 Fusion PCR with KanMX
Rpb11_RT_R	GGAAAACTCCATTTCAACG CC	Reverse Primer for rpb11 RT-PCR
Rpb11_RT_F	AATCAGCCAGAACGGTAT GAGC	Forward Primer for rpb11 RT-PCR
Rpb1_m1_f	GGCTGCCCAAAGTACAAT ACGTAAAGATGGCC	Forward Mutagenic Primer for rpb1-P180S ts
Rpb1_m1_r	GGCCATCTTTACGTATTGT ACTTTGGGCAGCC	Reverse Mutagenic Primer for rpb1-P180S ts
Rpb2_m902_f	GGCGAAGATATTATTATTG AGAAAACAGCTCCAATTCC CC	Forward Mutagenic Primer for rpb2-G902D ts
Rpb2_m902_r	GGGGAATTGGAGCTGTTTT CTCAATAATAATATCTTCG CC	Reverse Mutagenic Primer for rpb2-G902D ts
Rpb2_m1057_f	GGCTTTCAGTCGAGAGACT TTGAAGTTATGTACCATGG	Forward Mutagenic Primer for rpb2-G1057E ts

Rpb2_m1057_r	CCATGGTACATAACTTCAAAGTCTCTCGACTGAAAGCC	Reverse Mutagenic Primer for rpb2-G1057E ts
Rpb3_m1_f	GGTCACCTACTTCACCTGTTGCATTTGAATACG	Forward Mutagenic Primer for rpb3-A176P ts
Rpb3_m1_r	CGTATTCAAATGCAACAGGTGAAGTAGGTGACC	Reverse Mutagenic Primer for rpb3-A176P ts
Rpb1_f	AAAGCCAGTTTTCCACATCG	Cloning and transformation for rpb1-P180S ts
Rpb1_r	CAGAGACAATATGCGCAGGA	Cloning and transformation for rpb1-P180S ts
Rpb2_f	CGATGATTCTTGGCATCCTT	Cloning and transformation for rpb2 ts
Rpb2_r	GGCAATCACGTTCCATTTCT	Cloning and transformation for rpb2 ts
Rpb3_f	AGCTCCAACATCGATGAACC	Cloning and transformation for rpb3-A176P ts
Rpb3_r	AACCGACACTTTCGACATCC	Cloning and transformation for rpb3-A176P ts
Rpb1_seq_f	AACTATGGCGGATTGTCCTG	Sequencing rpb1
Rpb1_seq_r	CCCGCTATTTCATTGTCCAT	Sequencing rpb1
Rpb2_seq_f	AGCATCACGTCAAATGCAAG	Sequencing rpb2
Rpb2_seq_r	CTCGGTACGCATCAGAACAA	Sequencing rpb2
Rpb3_seq_f	TGCAGAGATTCCAACAGTGG	Sequencing rpb3
Rpb3_seq_r	ACCACGTGTTTTCTTCACCA	Sequencing rpb3
DirFus_ADH1_Term_R	GGTGAAGAAAACACGTGGTAATTACGCTGCTCAGTGCTGA	Rpb3 Fusion PCR with KanMX
DirFus_ Rpb3_Gene_F	TCAGCACTGAGCAGCGTAATTACCACGTGTTTTCTTCACC	Rpb3 Fusion PCR with KanMX
Rpb3_Nest_Gene	AAACTCGGTGGATTTTGTCC	Rpb3 Fusion PCR with KanMX
Rpb3_mA131T_f	GAAAGAGTTGCAATCTCGCTGCGACGTG	Forward Mutagenic Primer for rpb3-A131T ts
Rpb3_mA131T_r	CACGTCGCAGCGAGATTGCAACTCTTTC	Reverse Mutagenic Primer for rpb3-A131T ts
Rpb3_mA246G_f	CCTAGTTGAAATTAATGTGAGTACCTCAGTCATGCC	Forward Mutagenic Primer for rpb3-A246G ts
Rpb3_mA246G_r	GGCATGACTGAGGTACTCACATTAATTTCAACTAGG	Reverse Mutagenic Primer for rpb3-A246G ts
Rpb3_mT339C_f	CCGCCTCCAGTAGGTTCGGAATATACGCGC	Forward Mutagenic Primer for rpb3-T339C ts

Rpb3_mT339C_r	GCGCGTATATTCCGAACCTACTGGAGGCGG	Reverse Mutagenic Primer for rpb3-T339C ts
Rpb3_mT775C_f	GGAACCGCGTCGGTTCTATATGGATGTCG	Forward Mutagenic Primer for rpb3-T775C ts
Rpb3_mT775C_r	CGACATCCATATAGAACCGACGCGGTTCC	Reverse Mutagenic Primer for rpb3-T775C ts
*ssa2*_Up_Fusion_R	GATGGTGATGATGGTGATGCATATTGATGTGAAAAGACAAAAAAC	*ssa2* C-term HA tag Fusion PCR
*ssa2*_Tag_Fusion_F	GTTTTTTGTCTTTTCACATCAATATGCATCACCATCATCACCATC	*ssa2* C-term HA tag Fusion PCR
*ssa2*_Tag_Fusion_R	ATTCCGATTGATTTGCTCATACCACCCCCGCCCGATAACTT	*ssa2* C-term HA tag Fusion PCR
*ssa2*_Gene_Fusion_F	AAGTTATCGGGCGGGGGTGGTATGAGCAAATCAATCGGAAT	*ssa2* C-term HA tag Fusion PCR
*ssa2*_Gene_Nest_R	GTCACCAATGAGACGCTCGGT	*ssa2* C-term HA tag Fusion PCR
*ssa2*_N-Tag_Long_F	ATTTATATCCCACGTTCCTTTTTATAATCCCCAAATAAACTTTAGTCTTTTTTATTTGTTTTTTGTCTTTTCACATCAATATGCATCACCATCATCACCATC	Insertion of *ssa2* N-term HA tag
*ssa2*_N-Tag_Long_R	ATAATTTCGACACGGTTGTTGGAAAAGTGTCCAACACAAGAGTAGGTAGTACCCAAATCAATTCCGATTGATTTGCTCATACCACCCCCGCCCGATAACTT	Insertion of *ssa2* N-term HA tag
*ssa2*_HA_int_R	GGCGTAATCTGGAACATCGT	Testing integration at N-Term of *ssa2*
*ssa2*_HA_int_F	CGATGTTCCAGATTACGCCCA	Testing integration at N-term of *ssa2*
*ssa2*_nmt1_int_F	GCAATGTGCAGCGAAACTAA	Testing integration at N-term of *ssa2*
*ssa2*_seq_F	CACGCTCTTCCAGGACTTTC	Sequencing *ssa2-95*
*ssa2*_comp_F	ATTATCGCCAACGATCAAGG	*ssa2-95* Complimentation Fragment
*ssa2*_comp_R	CAAGGTAGGCCTCAGCAGTC	*ssa2-95* Complimentation Fragment
*ssa2*_seq_R	TTCAAACCGGCAATAAGACC	Sequencing *ssa2-95*
*ssa2*_spz_F	ATATCCCACCCTCCCAACAT	*ssa2* upstream primer
*ssa2*_Tag_Long_F	CCGCTCCTGGTGCTGCTCCTGGTGCTGCTCCTGGTGCTGCCCCTGGTGGTGACAATGGTCCTGAGGTTGAGGAGGTTGATCGGATCCCCGGGT	Insertion of *ssa2* C-term HA tag

	TAATTAA	
*ssa2*_Tag_Long_R	ATTACAAAAAAACATTGTTCACAACATAATTTTCACAATTAAAACCGCCCATTAAAAGATTTTTCTTCATGCACAGTATTGAATTCGAGCTCGTTTAAAC	Insertion of *ssa2* C-term HA tag
*ssa2*_Del_Long_F	ATTTATATCCCACGTTCCTTTTTATAATCCCCAAATAAACTTTAGTCTTTTTTATTTGTTTTTTGTCTTTTCACATCAATCGGATCCCCGGGTTAATTAA	Replacement of *ssa2* with KanMX
*ssa2*_int_F	GCTAACCCCATCATGGCTAA	Testing integration at C-term of *ssa2*
*ssa2*_int_R	CAAAGCAAACGAAAACACGA	Testing integration at C-term of *ssa2*
Bioneer_cp-N1	CGTCTGTGAGGGGAGCGTTT	Testing Bioneer *ssa1*Δ
Bioneer_cp-C3	GGCTGGCCTGTTGAACAAGTCTGGA	Testing Bioneer *ssa1*Δ
ssa1_Term_R	ATTGCATTCGTGGTTGGTTC	Testing Bioneer *ssa1*Δ
ssa1_Prom_F	CACGTCCAGCAAATTCTCAA	Testing Bioneer *ssa1*Δ
sgo2_C-term_F	CAAAACAGAAACCGCAAACA	Checking C-term Flag on *sgo2*$^{+}$
sgo2_Down_316bp_R	AGGGCGATCGAGGTACACTA	Checking C-term Flag on *sgo2*$^{+}$
TFIID_fwd	TGCTTTACCCACCACGGCCTCGCAAG	Splicing Assay for TFIID mRNA
TFIID_rev	TTCTGCATTACGTGCATGTAGCGC	Splicing Assay for TFIID mRNA
CDC2_fwd	GCTAGTGAACGGTGTAAATTTTTGC	Splicing Assay for CDC2 mRNA
CDC2_rev	CCCCAGTGGAATAATGTCTTGATC	Splicing Assay for CDC2 mRNA
ade9_F	CAGTCGTATGGTACTTCAAGTTCTG	Splicing Assay for *ade9*$^{+}$ mRNA
ade9_R	AAGCAAGGCAGGCATTTGGATTG	Splicing Assay for *ade9*$^{+}$ mRNA
Fbp A	AATGACAATTCCCCACTAGCC	Fbp loading control for ChIP
Fbp B	ACTTCAGCTAGGATTCACCTGG	Fbp loading control for ChIP
Cen F	GAAAACACATCGTTGTCTTCAGAG	otr primer for RT-PCR, ChIP
Cen R	CGTCTTGTAGCTGCATGTGAA	otr primer for RT-PCR, ChIP

Act F	GGCATCACACTTTCTACAACG	Actin loading control for RT-PCR
Act R	GAGTCCAAGACGATACCAGTG	Actin loading control for RT-PCR
q_cen(dg)_FOR	AATTGTGGTGGTGTGGTAATAC	dg for qPCR of ChIP
q_cen(dg)_REV	GGGTTCATCGTTTCCATTCAG	dg for qPCR of ChIP
q_act_FOR	GGTTTCGCTGGAGATGATG	Euchromatin control for qPCR of ChIP
q_act_REV	ATACCACGCTTGCTTTGAG	Euchromatin control for qPCR of ChIP
IK5	CTAGTCTAGTCTAGTCTAGTCTAGT	SnoRNA58 Probe for Northern
IK7	GATGAAATTCAGAAGTCTAGCATC	SnoRNA58 Probe for Northern
IK8	ATTCCTTTCTGAACCTCTCTGTTAT	Centromere siRNA Probe for Northern
IK9	TTTGATGCCCATGTTCATTCCACTTG	Centromere siRNA Probe for Northern
IK10	GGGAGTACATCATTCCTACTTCGATA	Centromere siRNA Probe for Northern

2.9 Strains

Strain Number	Genotype	Source
1180	h+ *ade6$^+$-210 leu1-32 ura4-D18* otr1R (dg-glu) Sph1::*ade6$^+$*	
1181	h- *ade6$^+$-210 leu1-32 ura4-D18* otr1R (dg-glu) Sph1::*ade6$^+$*	
A3418	h+ otrIR SphI::*ade6$^+$ ura4-DSE/D18 leu1-32 rpb3-1A(S55R)*-Kan	This Study
A3419	h+ otrIR SphI::*ade6$^+$ ura4-DSE/D18 leu1-32 rpb3-1B(G132R)*-Kan	This Study
A3420	h+ otrIR SphI::*ade6$^+$ ura4-DSE/D18 leu1-32 rpb3-6E(H158R)*-Kan	This Study
A3421	h+ otrIR SphI::*ade6$^+$ ura4-DSE/D18 leu1-32 rpb3-7E(R147C, R205W)*-Kan	This Study
A3422	h+ otrIR SphI::*ade6$^+$ ura4-DSE/D18 leu1-32 rpb3-8E(S100L)*-Kan	This Study
A3424	h+ otrIR SphI::*ade6$^+$ ura4-DSE/D18 leu1-32 rpb3-11E(H158D)*-Kan	This Study
A3428	h+ otrIR SphI::*ade6$^+$ ura4-DSE/D18 leu1-32 rpb3-16E(E60A)*-Kan	This Study
A3434	h+ otrIR SphI::*ade6$^+$ ura4-DSE/D18 leu1-32 rpb3-20B(G132R)*-Kan	This Study
A3439	h+ otrIR SphI::*ade6$^+$ ura4-DSE/D18 leu1-32 rpb3-32E(H158R, P163P)*-Kan	This Study
A3442	h+ otrIR SphI::*ade6$^+$ ura4-DSE/D18 leu1-32 rpb3-33C*-Kan	This Study
7005	h+ *dcr1*::NATR SphI::*ade6$^+$* leu? *his? ade6$^+$-210 ura4-D18*	
7036	h- *dcr1*::NATR SphI::*ade6$^+$* leu? *his? ade6$^+$-210 ura4-D18*	
A4870	h? otrIR SphI::*ade6$^+$ ura4-DSE/D18 leu1-32 rpb11-A2*-Kan	This Study
A4871	h? otrIR SphI::*ade6$^+$ ura4-DSE/D18 leu1-32 rpb11-A3*-Kan	This Study
A4872	h? otrIR SphI::*ade6$^+$ ura4-DSE/D18 leu1-32 rpb11-A7*-Kan	This Study
A4873	h? otrIR SphI::*ade6$^+$ ura4-DSE/D18 leu1-32 rpb11-A9*-Kan	This Study
A4874	h? otrIR SphI::*ade6$^+$ ura4-DSE/D18 leu1-32 rpb11-A10(K61E)*-Kan	This Study
A4875	h? otrIR SphI::*ade6$^+$ ura4-DSE/D18 leu1-32 rpb11-A11(E21D)*-Kan	This Study
A4876	h? otrIR SphI::*ade6$^+$ ura4-DSE/D18 leu1-32 rpb11-A12(W108stop)*-Kan	This Study
A4877	h? otrIR SphI::*ade6$^+$ ura4-DSE/D18 leu1-32 rpb11-A13(T40I)*-Kan	This Study

A4893	h? otrIR SphI::*ade6*$^{+}$ *ura4-DSE/D18 leu1-32 rpb3-1A(S55R)*-Kan	This Study
A4894	h? otrIR SphI::*ade6*$^{+}$ *ura4-DSE/D18 leu1-32 rpb3-1B(G132R)*-Kan	This Study
A4895	h? otrIR SphI::*ade6*$^{+}$ *ura4-DSE/D18 leu1-32 rpb3-6E(H158R)*-Kan	This Study
A4896	h? otrIR SphI::*ade6*$^{+}$ *ura4-DSE/D18 leu1-32 rpb3-8E(S100L)*-Kan	This Study
A4897	h? otrIR SphI::*ade6*$^{+}$ *ura4-DSE/D18 leu1-32 rpb3-11E(H158D)*-Kan	This Study
A4898	h? otrIR SphI::*ade6*$^{+}$ *ura4-DSE/D18 leu1-32 rpb3-16E(E60A)*-Kan	This Study
A4899	h? otrIR SphI::*ade6*$^{+}$ *ura4-DSE/D18 leu1-32 rpb3-20B(G132R)*-Kan	This Study
A4900	h? otrIR SphI::*ade6*$^{+}$ *ura4-DSE/D18 leu1-32 rpb3-32E(H158R, P163P)*-Kan	This Study
A5062	h? otrIR SphI::*ade6*$^{+}$ *ura4-DSE/D18 leu1-32 rpb3-1A(S55R)*-Kan	This Study
A5063	h? otrIR SphI::*ade6*$^{+}$ *ura4-DSE/D18 leu1-32 rpb3-1B(G132R)*-Kan	This Study
A5064	h? otrIR SphI::*ade6*$^{+}$ *ura4-DSE/D18 leu1-32 rpb3-6E(H158R)*-Kan	This Study
A5065	h? otrIR SphI::*ade6*$^{+}$ *ura4-DSE/D18 leu1-32 rpb3-8E(S100L)*-Kan	This Study
A5066	h? otrIR SphI::*ade6*$^{+}$ *ura4-DSE/D18 leu1-32 rpb3-11E(H158D)*-Kan	This Study
A5067	h? otrIR SphI::*ade6*$^{+}$ *ura4-DSE/D18 leu1-32 rpb3-16E(E60A)*-Kan	This Study
A5068	h? otrIR SphI::*ade6*$^{+}$ *ura4-DSE/D18 leu1-32 rpb3-20B(G132R)*-Kan	This Study
A5069	h? otrIR SphI::*ade6*$^{+}$ *ura4-DSE/D18 leu1-32 rpb3-32E(H158R, P163P)*-Kan	This Study
A5229	h? otrIR SphI::*ade6*$^{+}$ *ura4-DSE/D18 leu1-32 rpb11-A2*-Kan	This Study
A5230	h? otrIR SphI::*ade6*$^{+}$ *ura4-DSE/D18 leu1-32 rpb11-A3*-Kan	This Study
A5231	h? otrIR SphI::*ade6*$^{+}$ *ura4-DSE/D18 leu1-32 rpb11-A7*-Kan	This Study
A5232	h? otrIR SphI::*ade6*$^{+}$ *ura4-DSE/D18 leu1-32 rpb11-A9*-Kan	This Study
A5233	h? otrIR SphI::*ade6*$^{+}$ *ura4-DSE/D18 leu1-32 rpb11-A10(K61E)*-Kan	This Study
A5234	h? otrIR SphI::*ade6*$^{+}$ *ura4-DSE/D18 leu1-32 rpb11-A11(E21D)*-Kan	This Study
A5235	h? otrIR SphI::*ade6*$^{+}$ *ura4-DSE/D18 leu1-32 rpb11-A13(T40I)*-Kan	This Study
A5522	h+ otrIR SphI::*ade6*$^{+}$ *ura4-DSE/D18 leu1-32 rpb3-1A(S55R)*-Kan	This Study
A5523	h+ otrIR SphI::*ade6*$^{+}$ *ura4-DSE/D18 leu1-32 rpb3-1B(G132R)*-Kan	This Study

A5524	h+ otrIR SphI::*ade6*$^{+}$ *ura4-DSE/D18 leu1-32 rpb3-6E(H158R)*-Kan	This Study
A5525	h+ otrIR SphI::*ade6*$^{+}$ *ura4-DSE/D18 leu1-32 rpb3-8E(S100L)*-Kan	This Study
A5526	h+ otrIR SphI::*ade6*$^{+}$ *ura4-DSE/D18 leu1-32 rpb3-11E(H158D)*-Kan	This Study
A5527	h+ otrIR SphI::*ade6*$^{+}$ *ura4-DSE/D18 leu1-32 rpb3-16E(E60A)*-Kan	This Study
A5528	h+ otrIR SphI::*ade6*$^{+}$ *ura4-DSE/D18 leu1-32 rpb3-20B(G132R)*-Kan	This Study
A5529	h+ otrIR SphI::*ade6*$^{+}$ *ura4-DSE/D18 leu1-32 rpb3-32E(H158R, P163P)*-Kan	This Study
A5725	h+ otrIR SphI::*ade6*$^{+}$ *ura4-DSE/D18 leu1-32 rpb11-A10(K61E)*-Kan	This Study
A5726	h+ otrIR SphI::*ade6*$^{+}$ *ura4-DSE/D18 leu1-32 rpb11-A11(E21D)*-Kan	This Study
A5727	h+ otrIR SphI::*ade6*$^{+}$ *ura4-DSE/D18 leu1-32 rpb11-A12(W108stop)*-Kan	This Study
A5728	h+ otrIR SphI::*ade6*$^{+}$ *ura4-DSE/D18 leu1-32 rpb11-A13(T40I)*-Kan	This Study
A5620	h? otrIR SphI::*ura4+ ura4-DSE/D18? leu1-32 ade6*$^{+}$*-210 his1? rpb3-1A*-Kan	This Study
A5622	h? otrIR SphI::*ura4+ ura4-DSE/D18? leu1-32 ade6*$^{+}$*-210 his1? rpb3-1B*-Kan	This Study
A5613	h? otrIR SphI::*ura4+ ura4-DSE/D18? leu1-32 ade6*$^{+}$*-210 his1? rpb3-8E*-Kan	This Study
A5615	h? otrIR SphI::*ura4+ ura4-DSE/D18? leu1-32 ade6*$^{+}$*-210 his1? rpb3-11E*-Kan	This Study
A5617	h? otrIR SphI::*ura4+ ura4-DSE/D18? leu1-32 ade6*$^{+}$*-210 his1? rpb3-16E*-Kan	This Study
A5889	h? otrIR SphI::*ura4+ ura4-DSE/D18? leu1-32 ade6*$^{+}$*-210 his1? rpb11-A10*-Kan	This Study
A5890	h? otrIR SphI::*ura4+ ura4-DSE/D18? leu1-32 ade6*$^{+}$*-210 his1? rpb11-A11*-Kan	This Study
A5891	h? otrIR SphI::*ura4+ ura4-DSE/D18? leu1-32 ade6*$^{+}$*-210 his1? rpb11-A12*-Kan	This Study
A5892	h? otrIR SphI::*ura4+ ura4-DSE/D18? leu1-32 ade6*$^{+}$*-210 his1? rpb11-A13*-Kan	This Study
1624	h- *ade6*$^{+}$*-210 leu1-32 his1-102 ura4-DS/E* otr1R-Sph1::*ura4+*	
A117	h+ *dcr1*::G418R otr1R(Sph1):*ura4+ ura4DS/E ade6*$^{+}$*-210 leu1-32 his3D*1	
6084	h- clr4::LEU2+ otr1Rsph1::*ura4+ Ura4DS/E arg3D*3 *leu1-32*	
7329	h- *rpb2-m203* otr1(sph):*ura4+ ade6*$^{+}$*-210 ura4-D18*	Mwakami Lab
A6705	h? *rpb11-A13*-Kan otrIR SphI::*ura4+ ura4-DSE/D18? leu1-32 ade6*$^{+}$*-210 his1?*	This Study
A5898	h? otrIR SphI::*ade6*$^{+}$ mat3-M::*ura4+ ura4-DSE/D18 leu1-32 ade6*$^{+}$*-210*/216 *rpb3-1A*-Kan	This Study

A5900	h? otrIR SphI::*ade6*$^+$ mat3-M::*ura4+ ura4-DSE/D18 leu1-32 ade6*$^+$*-210*/216 *rpb3-11E*-Kan	This Study
A5901	h? otrIR SphI::*ade6*$^+$ mat3-M::*ura4+ ura4-DSE/D18 leu1-32 ade6*$^+$*-210*/216 *rpb3-16E*-Kan	This Study
A5902	h? otrIR SphI::*ade6*$^+$ mat3-M::*ura4+ ura4-DSE/D18 leu1-32 ade6*$^+$*-210*/216 *rpb11-A10*-Kan	This Study
A5903	h? otrIR SphI::*ade6*$^+$ mat3-M::*ura4+ ura4-DSE/D18 leu1-32 ade6*$^+$*-210*/216 *rpb11-A11*-Kan	This Study
A5904	h? otrIR SphI::*ade6*$^+$ mat3-M::*ura4+ ura4-DSE/D18 leu1-32 ade6*$^+$*-210*/216 *rpb11-A12*-Kan	This Study
A5905	h? otrIR SphI::*ade6*$^+$ mat3-M::*ura4+ ura4-DSE/D18 leu1-32 ade6*$^+$*-210*/216 *rpb11-A13*-Kan	This Study
511	h90 mat3-M::*ura4+ ade6*$^+$*-216 leu1-32 ura4-D18*	
9299	h90 mat3::*ura4+ ura4DS/E dcr1*::NAT	
7429	h? mat3::*ura4+ clr4::leu2*	
A6077	h? *ade6*$^+$*-210 arg3-D4 his3-D1 leu1-32 ura4-D18* TM1::*arg3* tel::*his3 rpb3-1A*-Kan	This Study
A6078	h? *ade6*$^+$*-210 arg3-D4 his3-D1 leu1-32 ura4-D18* TM1::*arg3* tel::*his3 rpb3-11E*-Kan	This Study
A6079	h? *ade6*$^+$*-210 arg3-D4 his3-D1 leu1-32 ura4-D18* TM1::*arg3* tel::*his3 rpb3-16E*-Kan	This Study
A6080	h? *ade6*$^+$*-210 arg3-D4 his3-D1 leu1-32 ura4-D18* TM1::*arg3* tel::*his3 rpb11-A10*-Kan	This Study
A6081	h? *ade6*$^+$*-210 arg3-D4 his3-D1 leu1-32 ura4-D18* TM1::*arg3* tel::*his3 rpb11-A11*-Kan	This Study
A6082	h? *ade6*$^+$*-210 arg3-D4 his3-D1 leu1-32 ura4-D18* TM1::*arg3* tel::*his3 rpb11-A12*-Kan	This Study
A6083	h? *ade6*$^+$*-210 arg3-D4 his3-D1 leu1-32 ura4-D18* TM1::*arg3* tel::*his3 rpb11-A13*-Kan	This Study
3027	h+ *ade6*$^+$*-210 arg3-D3 his3*-D1 *leu1-32 ura4-D18/DS-E* TM1::*arg3* TM3::*ade6*$^+$ tel::*his3* otr2::*ura4+*	
A5382	h? *dcr1*::Nat *ura4-D18*/DS-E *leu1-32 ade6*$^+$*-210* arg3/D3 *his3*-D1 tel::*his3*	
2506	h? rik1::LEU2 *ade6*$^+$*-210 arg3-D4 his3-D1 leu1-32 ura4-DS/E* otr1RSph1::*ade6*$^+$ TEL1L-*his3*	
A386	h? *ade6*$^+$*-210 ura4-D18*/DS-E *dcr1*::KAN TM1:*arg3*	
3610	h? *ade6*$^+$*-210 arg3D*1 *his3D*1 *leu1-32 ura4D18 rik1::leu2* TM1::*arg3* otr2::*ura4+ his3*::tel	
2017	h+ (Xba1-Spe1)clr4::LEU2 *ade6*$^+$*-210 leu1-32 ura4-D18*	
5806	h? ago1::Kan otrIR SphI::*ade6*$^+$ *ura4-DSE/D18 leu1-32 ade6*$^+$*-210*	
A6697	h? *rpb3-11E*-Kan ago1::*ura4+* otrIR SphI::*ade6*$^+$ *ura4-DSE/D18 leu1-32 ade6*$^+$*-210*	This Study
A6700	h? *rpb3-11E*-Kan ago1::*ura4+* otrIR SphI::*ade6*$^+$ *ura4-DSE/D18 leu1-32 ade6*$^+$*-210*	This Study

A6701	h? *rpb11-A10*-Kan ago1::*ura4*+ otrIR SphI::*ade6*$^+$ *ura4-DSE/D18 leu1-32 ade6*$^+$*-210*	This Study
A6704	h? *rpb11-A10*-Kan ago1::*ura4*+ otrIR SphI::*ade6*$^+$ *ura4-DSE/D18 leu1-32 ade6*$^+$*-210*	This Study
A6689	h? *rpb3-11E*-Kan *dcr1*::Nat otrIR SphI::*ade6*$^+$ *ura4-DSE/D18 leu1-32 ade6*$^+$*-210*	This Study
A6691	h? *rpb3-11E*-Kan *dcr1*::Nat otrIR SphI::*ade6*$^+$ *ura4-DSE/D18 leu1-32 ade6*$^+$*-210*	This Study
A6693	h? *rpb11-A10*-Kan *dcr1*::Nat otrIR SphI::*ade6*$^+$ *ura4-DSE/D18 leu1-32 ade6*$^+$*-210*	This Study
A6695	h? *rpb11-A10*-Kan *dcr1*::Nat otrIR SphI::*ade6*$^+$ *ura4-DSE/D18 leu1-32 ade6*$^+$*-210*	This Study
2268	h? clr4::*ura4*+ otrIR SphI::*ade6*$^+$ *ura4-D18 leu1-32 ade6*$^+$*-210*	
A7096	h? *rpb3-11E*-Kan clr4::*ura4*+ otrIR SphI::*ade6*$^+$ *ura4-DSE/D18 leu1-32 ade6*$^+$*-210*	This Study
A7102	h? *rpb11-A10*-Kan clr4::*ura4*+ otrIR SphI::*ade6*$^+$ *ura4-DSE/D18 leu1-32 ade6*$^+$*-210*	This Study
A6102	h90 *rpb2-m203*-Kan otr1(sph)::*ade6*$^+$ *ade6*$^+$*-210 ura4-D18* leu1$^+$*/leu1-32*	This Study
A6746	h? *rpb2-m203*-Nat ago1::Kan otrIR SphI::*ade6*$^+$ *ura4-DSE/D18 leu1-32 ade6*$^+$*-210*	This Study
A6742	h? *rpb2-m203*-Kan *dcr1*::Nat otrIR SphI::*ade6*$^+$ *ura4-DSE/D18 leu1-32 ade6*$^+$*-210*	This Study
A7098	h? *rpb2-m203*-Kan clr4::*ura4*+ otrIR SphI::*ade6*$^+$ *ura4-DSE/D18 leu1-32 ade6*$^+$*-210*	This Study
707	h- *clr4-s5 ade6*$^+$*-210 leu1-32 ura4-DS/E* otr1R dg-glu (BamHI-Spe1) Sph1::*ura4*+	
1382	h+ *csp6-95 ade6*$^+$*-210 ura4-DS/E* or *D18* otr1R-Sph1::*ade6*$^+$	
1406	h+ *csp6-75 ade6*$^+$*-210 ura4-DS/E* or *D18* otr1R-Sph1::*ade6*$^+$	
7466	h+ *prp1-1*ts *ade6*$^+$::KAN otr1R sph::*ade6*$^+$ *lys1*::NAT *leu*? *ura4?*	
A2137	h? TS otr1R(sph1):*ade6*$^+$ *lys1*::NAT *ade6*$^+$::G418 *leu1-32 ura4-D18 csp6-95*	This Study
A3486	h+ otr1R(Sph1):*ade6*$^+$ *ade6*$^+$*-210 leu1-32 ura4-D18 csp6-95*	This Study
A3484	h+ otr1R(Sph1):*ade6*$^+$ *ade6*$^+$*-210 leu1-32 ura4-D18 csp6-75*	This Study
A2135	h? TS otr1R(sph1):*ade6*$^+$ *lys1*::NAT *ade6*$^+$::G418 *leu1-32 ura4-D18 csp1*	This Study
A2136	h? TS otr1R(sph1):*ade6*$^+$ *lys1*::NAT *ade6*$^+$::G418 *leu1-32 ura4-D18 csp2*	This Study
A5277	h+ otr1R(Sph1):*ade6*$^+$ *ade6*$^+$*-210 leu1-32 ura4-D18 ssa2-95*	This Study
A5619	h+ otr1R(Sph1):*ade6*$^+$ *ade6*$^+$*-210 leu1-32 ura4-D18 ssa2*+	This Study
A6624	h+ *ssa2*::Nat otrIR SphI::*ade6*$^+$ *ura4-DSE/D18 leu1-32 ade6*$^+$*-210*	This Study

A6683	h? *ssa2*::Nat otrIR 3xFlag-Ago1-Kan ?SphI::*ade6*$^+$? *ura4-DSE/D18 leu1-32 ade6*$^+$*-210*	This Study
A6733	h? *ssa2-75*-HA-Nat 3xFlag-Ago1-Kan SphI::*ade6*$^+$ *ade6*$^+$*-210 arg3D his3D leu1-32 ura4-D18*	This Study
A6618	h+ *ssa2*-HA-Nat otrIR SphI::*ade6*$^+$ *ura4-DSE/D18 leu1-32 ade6*$^+$*-210*	This Study
A6735	h? *ssa2*-HA-Nat 3xFlag-Ago1-Kan SphI::*ade6*$^+$ *ade6*$^+$*-210 arg3D his3D leu1-32 ura4-D18*	This Study
A6802	h? *ssa2-95*-HA-Nat 3xFlag-Ago1-Kan SphI::*ade6*$^+$ *ade6*$^+$*-210 arg3D his3D leu1-32 ura4-D18*	This Study
A6877	h? *ssa2*-HA-Nat 5xFlag-Clr4 *ade6*$^+$*-210 arg3D his3D leu1-32 ura4-D18/DSE* ?Kint2::*ura4*?	This Study
A6983	h? *ssa2-95*-HA-Nat 5xFlag-clr4 *ade6*$^+$*-210 arg3D his3D leu1-32 ura4-D18/DSE* ?Kint2::*ura4*?	This Study
A7008	h? *ssa2*-HA-Nat rdp1-3xFlag-Nat SphI::*ade6*$^+$ *ade6*$^+$*-210 arg3D his3D leu1-32 ura4-D18/DSE*	This Study
A7059	h? *ssa2-95*-HA-Nat rdp1-3xFlag-Nat SphI::*ade6*$^+$ *ade6*$^+$*-210 arg3D his3D leu1-32 ura4-D18/DSE*	This Study
A7384	h? *ssa2*-HA-Nat scm3-3xFlag-Kan *ade6*$^+$*-210 arg3D his3D leu1-32 ura4-D18/DSE*	This Study
A7556	h? *ssa2-95*-HA-Nat scm3-3xFlag-Kan *ade6*$^+$*-210 arg3D his3D leu1-32 ura4-D18/DSE*	This Study
A586	h? *eri1*::NATMX6 3xFlag-ago1-KANMX6	
A7269	h? *ssa2-95* eri1::Nat 3xFlag-Ago1-Kan ?SphI? *ura4-DSE/D18 leu1-32 ade6*$^+$*-210*	This Study
A6808	h? *nat-nmt1-ago1* otrIR SphI::*ade6*$^+$ *ura4-DSE/D18 leu1-32 ade6*$^+$*-210*	This Study
A6805	h+ *ssa2-95 nat-nmt1-ago1* otrIR SphI::*ade6*$^+$ *ura4-DSE/D18 leu1-32 ade6*$^+$*-210*	This Study
A6749	h+ *ssa2-75 nat-nmt1-ago1* otrIR SphI::*ade6*$^+$ *ura4-DSE/D18 leu1-32 ade6*$^+$*-210*	This Study

Chapter 3 - Generating RNAPII mutants

3.1. Introduction

In *S. pombe* the centromeres are flanked by repetitive DNA which is packed into heterochromatin. The heterochromatin is required for cohesion between the sister chromatids during cell division and loss of this pericentromeric heterochromatin causes missegregation of chromosomes. Marker genes inserted into heterochromatin are silenced by spreading of the heterochromatin over the gene. In *S. pombe* one of the main roles of the RNAi pathway is establish and maintain pericentromeric heterochromatin. Short-interfering (si)RNAs are processed, by the endonuclease Dicer, from long non-coding RNA transcribed from the pericentromeric repeats. The siRNA is taken up by the Argonaute protein and used to target the main RNAi effector complex, RITS, to cause heterochromatin to be formed over the regions homologous to the siRNA. Clr4 is the histone methyltransferase responsible for methylating histone H3 on lysine 9 (H3K9) which is an important heterochromatin modification and it is recruited to the pericentromere by RNAi factors.

Paradoxically RNAi-mediated heterochromatin formation is dependent on transcription of the target locus in order to generate the transcript which is diced into siRNAs. RNA polymerase II (RNAPII) is the RNA polymerase enzyme that transcribes the pericentromeric repeats which are processed into the siRNAs required for heterochromatin formation (Djupedal, 2005). Consistent with this, mutations in two of the subunits of RNAPII have been found to disrupt heterochromatic formation at the pericentromere. However, these two mutants appear to impair silencing in different ways, suggesting that the role of RNAPII in heterochromatin assembly may be complex.

Djupedal (2005), characterised the *rpb7-G150D* mutant found in the csp screen for mutants with centromeric silencing defects (Ekwall, 1999). The *rpb7-G150D* mutant has a two-fold decrease in general transcription, but it was shown that the centromeric transcripts were more specifically knocked down to much lower levels. The low levels of centromeric transcripts abolished the RNAi-dependent heterochromatin at the centromere, resulting in the observed de-

repression of the marker genes within the region. The phenotypes of the *rpb7-G150D* mutant therefore implied that RNAPII is important for heterochromatin formation at centromeres simply because it is required to transcribe the non-coding RNAs.

Characterisation of a mutation in the Rpb2 subunit (Kato, 2005) suggested that RNAPII may play a central role in mediating RNAi-directed heterochromatin formation. This mutant, *rpb2-m203*, was found to have normal levels of transcription, including transcription of the pericentromeric repeats, but had a defect in their processing into siRNA. Maintenance of heterochromatin at the pericentromere is completely dependent on the RNAi pathway whereas redundant pathways at the mating-type locus and the telomeres mean that RNAi is dispensable for heterochromatin maintenance at these loci. As with the *rpb7-G150D* mutant the silencing defect observed was specific to the pericentromeric heterochromatin region, with normal marker gene silencing at the mating-type locus, suggesting a specific role for the Rpb2 subunit in the RNAi pathway downstream of transcription. No further observations have been published concerning the role of RNAPII in the RNAi pathway, so the way in which RNAPII is involved in RNAi (apart from transcription) remains to be determined.

One of the difficulties in working with the *rpb2-m203* and *rpb7-G150D* mutants is that their silencing defects variegate, likely due to fluctuations in heterochromatin integrity across the marker gene (Trewick et al., 2007). This means that within a colony individual cells have differing levels of expression of the marker gene. The level of variegation can be sensitive to molecular manipulations like transformation with DNA. For this reason it would be useful to have RNAPII mutants with defects in RNAi-directed heterochromatin formation which do not suffer variegation. Additional RNAPII mutants may also provide new clues to the role(s) of RNAPII in this pathway.

One of the genetic tools which is currently lacking in the *S. pombe* community is an RNAPII mutant which has a tight conditional phenotype (ie good growth at the permissive temperature and complete loss of transcriptional activity at the restrictive temperature). This deficiency has been highlighted

recently by the discovery of transcripts under the kinetochore (TUKs) by Dr Choi in the Allshire lab (Choi et al., 2011), a study for which a conditional RNAPII mutant would have been very useful. For this reason I aimed to produce new ts RNAPII mutants as well as mutants with defects in pericentromeric heterochromatin formation.

In this chapter I describe the generation of a new set of RNAPII mutants, with the aim of facilitating further dissection of the role of RNAPII in heterochromatin assembly: RNAPII mutants with silencing defects could be used to identify possible genetic and physical interactions between RNAPII subunits and RNAi proteins. I have used 3 main strategies to produce RNAPII mutants. Two of the strategies employed site-directed mutagenesis techniques to re-make known ts mutations but these were unsuccessful. I will describe the third, successful, approach of random mutagenesis screening.

3.2 Results

3.2.1 Random Mutagenesis of RNAPII subunits; *rpb3*$^+$, *rpb11*$^+$

In order to generate RNAPII mutants I used random mutagenesis of two of the subunits and subsequently screened for mutants with defects in silencing of a marker gene inserted in the heterochromatic pericentromere region and/or temperature sensitivity (ts). RNAPII subunits 3 and 11 were selected as good candidates for three reasons; (1) These two subunits along with Rpb2 assemble to produce an RNAPII subassembly complex - which is homologous to the prokaryotic $\alpha_2\beta$ complex (Kimura et al., 1997) - which means mutations are more likely to produce a ts phenotype (Mitobe, 1999). This is because instability in the subassembly complex subunits due to mutation is more likely to cause the whole RNAPII complex to be unstable at high temperatures. (2) There is a large gene-free region downstream of both the *rpb3*$^+$ and *rpb11*$^+$ genes which means that other gene promoters are unlikely to be inadvertently affected by the insertion of the KanMX marker gene (3) The genes are both <1000bp in size allowing the entire gene to be amplified in a single mutagenic PCR reaction and would be more easily stitched into a fusion construct.

Mutagenesis of the *rpb3*$^{+}$ and *rpb11*$^{+}$ genes was carried out using mutagenic PCR of the appropriate *rpb* gene using genomic template DNA, aiming for one to two mutations per gene. To aid selection of transformants, fusion constructs were made (Figure 3.1) with the G418 resistance gene (KanMX) flanked by the mutated *rpb3 or rpb11* gene and its equivalent genomic downstream sequence. The fusion constructs were transformed using electroporation into a tester strain with the *ade6*$^{+}$ gene inserted in the pericentromeric outer repeats of centromere one. The transformants were initially selected for G418 resistance then replica plated, in parallel, onto Phloxin B plates grown at 36°C to test temperature sensitivity and low adenine plates grown at 25°C to test for silencing defects. The screen for *rpb3* mutants was carried out first followed by a more comprehensive screen for *rpb11* mutants.

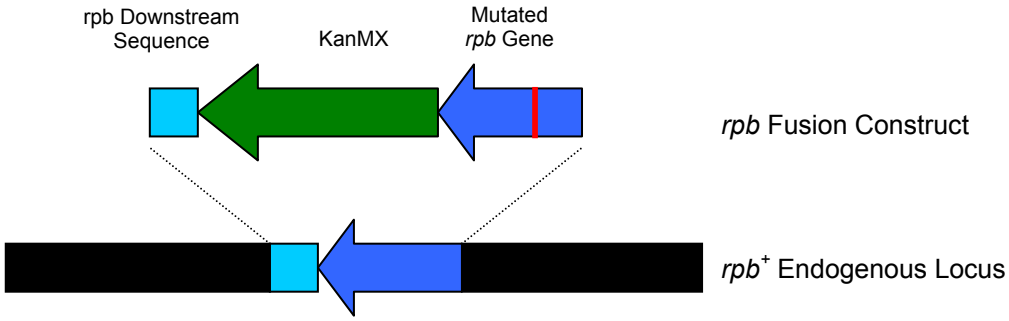


Figure 3.1 – An illustration of the strategy used to insert a mutated *rpb* gene to the genome. The construct is transformed into the cell and transformants carrying the construct are selected using the antibiotic G418, due to the presence of the G418 resistance gene KanMX. Integration into the genome can occur by replacing the endogenous gene through homologous recombination.

3.2.2 *rpb3* mutants isolated in the screen

For *rpb3* 35 potential mutants were picked from the initial screen presenting differing degrees of silencing defect (white to dark pink on low adenine plates) and/or temperature sensitivity (ts). Colour acts as readout of expression levels of the *ade6*$^{+}$ marker gene when cells are grown on low adenine plates. When the

colony is red then the *ade6*$^{+}$ gene is repressed by the heterochromatin; white colonies have defective heterochromatin at the pericentromere and thus express *ade6*$^{+}$. There were no mutants isolated that displayed a ts phenotype without a silencing defect. The mutants were tested for integration of the KanMX marker gene at the *rpb3* locus by PCR to ensure that the defects observed were not due to random integration. Comparative plating assays were performed to verify the silencing defects and temperature sensitivity of these mutants (Appendix – Figure 1). Cells in which centromere function is defective tend to be hypersensitive to the microtubule-destabilising drug Thiabendazol (TBZ) thus the mutants were tested for by spotting onto plates containing TBZ. The 10 mutants with the strongest phenotypes are summarised in Table 3.1.

Strain	Low adenine	No adenine	TBZ	Temperature sensitive
Wildtype	5	-	+++	no
*dcr1*Δ	1	+++	-	partially sensitive at 36°C
1A	2	+++	-	dead at 32°C
1B	2	+++	-	dead at 32°C
6E	1	+++	-	no
7E	2	+	+++	no
8E	3	++	+	dead at 36°C
11E	2	+++	+++	no
16E	2	+++	+++	no
20B	1	+++	-	dead at 32°C
32E	2	+++	+++	no
33C	2	+++	+	no

++	Good Growth	+	Intermediate Growth	-	No Growth

Table 3.1 – Summary of the *rpb3* mutants with the strongest phenotypes.
Expression of the *ade6*$^{+}$ gene causes colonies to have a white colour whereas repression of *ade6*$^{+}$ causes a red pigment to accumulate. The *ade6*$^{+}$ gene is located in the pericentromeric heterochromatin thus expression of the gene indicates defective silencing in the region. The colour on low adenine plates is scored 1 (white/defective silencing) to 5 (red/normal silencing). Growth on restrictive media plates has been given an arbitrary score ranging from '+++' for good growth to '- ' for a complete lack of growth. Growth on plates with no adenine is indicative of defective silencing. A defect in pericentromeric heterochromatin results in lagging chromosomes and sensitivity to the microtubule destabilising drug TBZ.

All of the mutants picked have reduced growth rates compared to wildtype strains. Three of the mutants with the slowest growth rates (*rpb3-1A, rpb3-1B* and *rpb3-20B*) also have the strongest silencing defects and are ts even at 32°C (generally a semi-permissive temperature). All of the other mutants display quite strong alleviation of *ade6*$^{+}$ repression but do not have strong TBZ sensitivity suggesting that in these cases pericentromeric heterochromatin is only partially defective. Similar to *rpb2-m203* and *rpb7-G150D*, all of the mutants have variegating silencing defects as indicated by a mix of red, pink or white colonies when a single colony is streaked out on a low adenine plate. The three ts mutants (*rpb3-1A, rpb3-1B* and *rpb3-20B*) have the lowest level of variegation forming only pink but mainly white colonies upon streaking to low adenine plates.

3.2.3 *rpb11* mutants isolated in the screen

For *rpb11* 317 potential mutants were picked presenting different phenotypes: 72 ts mutants, 258 white/pink colonies (possible defective silencing), 7 dark red colonies (possible enhancers of silencing). These mutants were spotted onto low adenine plates in 96-well format in order to quickly assess the strength of their silencing defects (Figure 3.2).

The 72 potential ts mutants were streaked to plates and the streaks were replica plated onto Phloxin B at 25°C and 36°C. It was found that only 3 mutants displayed a true ts phenotype (ie death at 36°C) whilst the majority of the strains were dark pink at both 25°C and 36°C but were healthy at both temperatures. A possible explanation for the number of false positives is that I was picking any colony that showed a dark staining on Phloxin B plates at 36°C. Diploids also stain darker than haploids on Phloxin B, so any diploid colonies present would also be picked. The three ts mutants and 13 of the white mutants were tested for silencing defects, temperature sensitivity and TBZ sensitivity as before, in order to assess and confirm the phenotypes (Appendix – Figure 2). The 10 mutants with the strongest phenotypes are summarised in Table 3.2.

Strain	Low Adenine	No Adenine	TBZ	Temperature Sensitivity
Wildtype	5	-	+++	no
dcr1Δ	1	+++	-	partially sensitive at 36°C
csp6-95	1	+++	-	Dead at 36°C
ts1	5	-	+++	Dead at 36°C
A2	3	++	+	no
A3	3	++	+	no
A6	1	+++	-	no
A7	1	+++	-	no
A9	2	+++	+	no
A10	1	+++	+	no
A11	2	+++	+	no
A12	1	+++	-	no
A13	1	+++	-	no

++	Good Growth	+	Intermediate Growth	-	No Growth

Table 3.2 – Summary of the *rpb11* mutants with the strongest phenotypes. The *ade6*$^{+}$ gene is positioned in the pericentromeric outer repeats. When heterochromatin silencing is fully functional the *ade6*$^{+}$ gene is repressed and so the colonies gain a red colour when grown on low adenine plates. Mutants defective in silencing have a white colour as the *ade6*$^{+}$ gene is expressed. Colour on low adenine plates is scored 1 (white/defective silencing) to 5 (red/wildtype silencing). Growth on restrictive media plates has been given an arbitrary score ranging from '+++' for good growth to '- ' for a complete lack of growth. Growth on plates with no adenine is indicative of defective silencing. Mutants with defective heterochromatin at the pericentromere also have defects in chromosome segregation leading to sensitivity to the microtubule destabilising drug TBZ.

All of the *rpb11* mutants summarised in Table 3.2 have reduced growth rates compared to wildtype strains although *rpb11-A13* is noticeably faster growing than the other mutants. *rpb11-ts1* displays a relatively good growth rate at the permissive temperature (25°C) and complete death at the restrictive temperature (36°C). However, although promising, *rpb11-ts1* was later ruled out as the phenotype did not co-segregate with the marker at the *rpb11* locus (Table 3.4). The *rpb11* mutants with silencing defects all display TBZ sensitivity suggesting that these mutants generally have more complete defects in heterochromatin compared to the majority of the *rpb3* mutants isolated (Table 3.1). As with the *rpb3* mutants the *rpb11* mutants with silencing defects have

variegating phenotypes where an individual colony can be made up of a mix of red, pink or white cells. This is a well known phenomenon in silencing mutants and is likely due to fluctuating/unstable heterochromatin (Trewick et al., 2007).

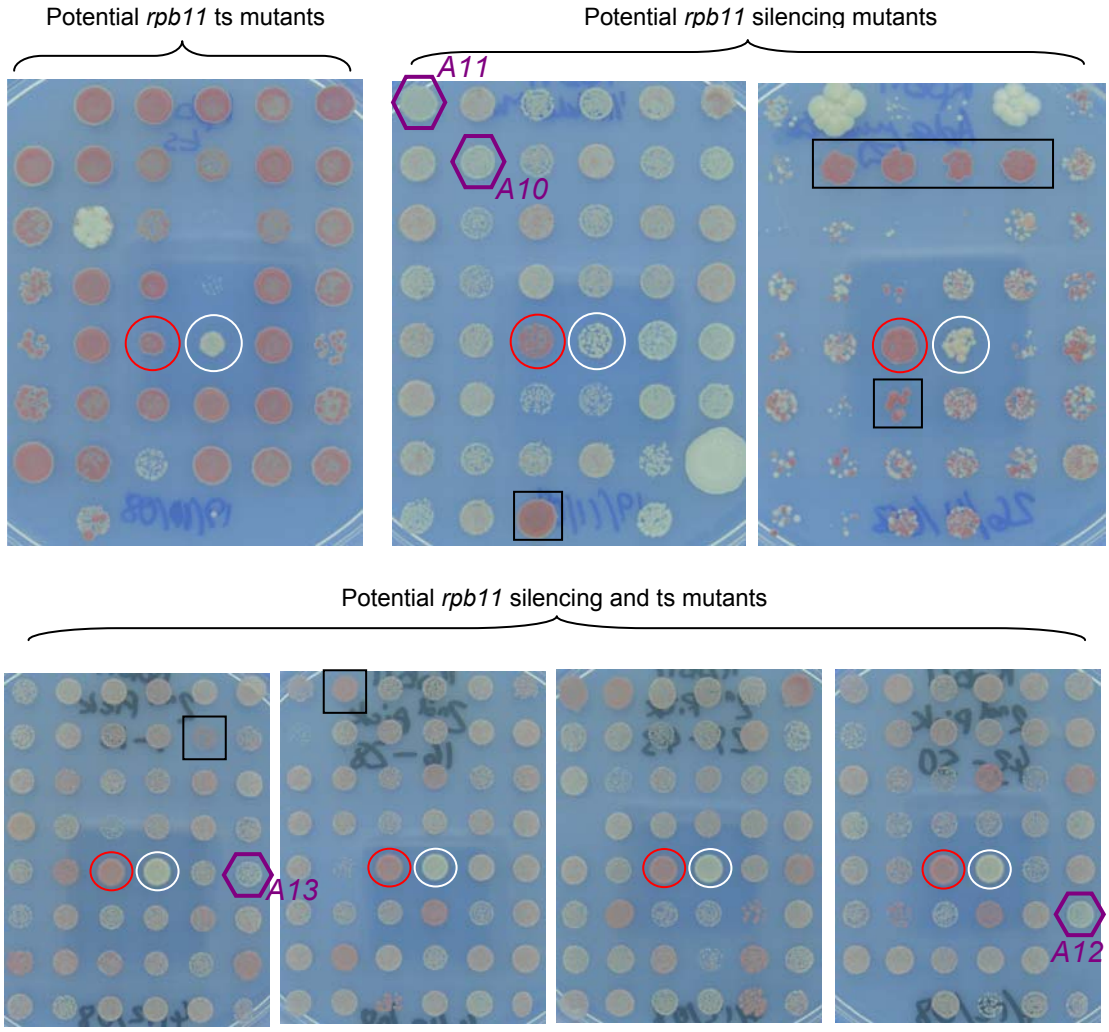


Figure 3.2 – Low adenine colour assay of potential *rpb11* mutants. The white colonies alleviate repression of the *ade6+* gene and thus have defective silencing in pericentromeric heterochromatin. Red colonies display normal silencing in pericentromeric heterochromatin as the *ade6+* gene is repressed. Dark red colonies may be due to enhanced silencing in pericentromeric heterochromatin.

3.2.4 Identifying mutated residues in select RNAPII mutants

PCR was used to test if the KanMX cassette had integrated at the correct locus for the 20 top mutants for Rpb3 and Rpb11. This confirmed that the KanMX gene was integrated at the correct locus in all strains tested. Sequencing of the relevant *rpb* gene from the mutants confirmed that mutations had been introduced at a rate of 1-2 mutations per gene, although some of the mutants had no mutations in the target gene (Table 3.3).

Strain	Mutations (Nucleotide)	Mutations (Amino Acid)
rpb11-ts1	Not sequenced due to backcross results (Table 3.4)	
rpb11-A2	----------No Mutation Found----------	
rpb11-A3	T224A	Intron 2
rpb11-A6	----------No Mutation Found----------	
rpb11-A7	----------No Mutation Found----------	
rpb11-A9	----------No Mutation Found----------	
rpb11-A10	A272G	K61E
rpb11-A11	G108T	E21D
rpb11-A12	G414A	W108Stop
rpb11-A13	G-51C, T83C, C164T	Promoter, Intron 1, T40I
rpb3-1A	T251C, T892C	S55R, P268P
rpb3-1B	*G512A*	G132R
rpb3-6E	*A591G*	*H158R*
rpb3-7E	C557T, C731T	R147C, R205W
rpb3-8E	C387T	S100L
rpb3-11E	*C590G*	*H158D*
rpb3-16E	A267C	E60A
rpb3-20B	*G512A*	G132R
rpb3-32E	*A591G, T607C*	*H158R, P163P*
rpb3-33C	----------No Mutation Found----------	

Table 3.3 – Sequencing results for the top 20 RNAPII mutants. Sanger sequencing was used to identify mutations in the rpb3 or rpb11 open reading frame. Two reactions were compared per mutant. The mutations highlighted by colours indicate that the codon mutated is also mutated in the other mutants highlighted with the same colour.

Mutants lacking any mutation in the relevant gene were ruled out from further analysis. The remaining mutants were backcrossed twice to the parental strain in order to confirm that the phenotype was linked to the G418 resistance

gene at the relevant *rpb* locus (Table 3.4). The mutant defect should co-segregate with the G418 resistance in 100% of the progeny because the KanMX has been shown to be present at the *rpb* gene of interest in each mutant. Silencing defects or ts phenotypes that are unlinked to G418 resistance are thus not caused by the *rpb* gene and these mutants can be discarded.

	Backcross 1			Backcross 2			
Mutant	**Colonies Counted**	**G418^R - Defective* Phenotype**	**G418^S - Wildtype Phenotype**	**Colonies Counted**	**G418^R - Defective* Phenotype**	**G418^S - Wildtype Phenotype**	**Linkage**
rpb11-ts1	25	60.0%	53.3%	N/A	N/A	N/A	Unlinked
rpb11-A2	25	88.9%	93.8%	25	100.0%	100.0%	Linked
rpb11-A3	25	85.7%	81.8%	25	100.0%	100.0%	Linked
rpb11-A6	25	54.5%	50.0%	N/A	N/A	N/A	Unlinked
rpb11-A7	25	85.7%	100.0%	25	100.0%	100.0%	Linked
rpb11-A9	25	83.3%	100.0%	25	100.0%	100.0%	Linked
rpb11-A10	25	88.9%	93.8%	25	100.0%	100.0%	Linked
rpb11-A11	25	100.0%	100.0%	25	100.0%	100.0%	Linked
rpb11-A12	25	90.0%	93.3%	25	100.0%	100.0%	Linked
rpb11-A13	25	80.0%	100.0%	25	100.0%	100.0%	Linked
rpb3-1A	25	100.0%	100.0%	25	100.0%	100.0%	Linked
rpb3-1B	14	100.0%	100.0%	25	100.0%	100.0%	Linked
rpb3-6E	25	100.0%	100.0%	25	100.0%	100.0%	Linked
rpb3-7E	25	50.0%	54.5%	N/A	N/A	N/A	Unlinked
rpb3-8E	25	100.0%	100.0%	25	87.5%	100.0%	Linked
rpb3-11E	25	85.7%	100.0%	25	93.3%	90.0%	Linked
rpb3-16E	25	100.0%	100.0%	25	93.3%	100.0%	Linked
rpb3-20B	4	100.0%	100.0%	25	100.0%	100.0%	Linked
rpb3-32E	25	81.3%	100.0%	25	93.3%	100.0%	Linked

Table 3.4 – Summary of backcross results for the new RNAPII mutants. In each case the mutant strain was crossed to the parental tester strain and the resulting progeny were tested for G418 resistance as well as defects in silencing and in some cases ts. G418 resistance should be linked to the defect phenotype in 100% of the progeny as the KanMX (G418 resistance gene) is known to be adjacent to the mutated *rpb* gene in each mutant. The mutants with silencing defects or temperature sensitivity which is unlinked to G418 resistance are highlighted in red.
* silencing defects and also ts in some cases

In three of the mutants, the *rpb* locus was found to be unlinked to the mutant phenotype - these were excluded from further analysis. These unlinked mutations presumably arose in other genes during transformation. Most of the

mutants did not display 100% co-segregation of the silencing/ts phenotypes with the antibiotic resistance. This is mainly due to the fact that the silencing phenotype in these mutants is variable so that a proportion of colonies appear to be wildtype. Importantly the mutants with the most obvious defects, such as ts and strong silencing defects, displayed 100% segregation of both phenotypes with G418 resistance.

In order to demonstrate that the mutation sequenced for each of the mutants is responsible for the observed phenotype, the mutations were regenerated using the mutant DNA as a template. The results are summarised in Table 3.5.

Mutant	Total Colonies (on -Ade)	White (on 1/10th Ade)	Pink (on 1/10th Ade)	Red (on 1/10th Ade)	ts	not ts
rpb3-1A*	30	14	0	16	14	16
rpb3-1B*	34	18	0	16	18	16
rpb3-6E	13	0	9	4	-	-
rpb3-8E*	20	0	12	8	12	8
rpb3-11E	32	20	0	12	-	-
rpb3-16E	26	0	13	13	-	-
rpb3-20B*	20	7	0	13	7	13
rpb3-32E	6	6	0	0	-	-
rpb11-A10	140	0	98	42	-	-
rpb11-A11	135	0	102	33	-	-
rpb11-A12	31	0	20	11	-	-
rpb11-A13	53	0	39	14	-	-

* denotes originally ts mutant

Table 3.5 – Phenotypes of G418 resistant colonies after transformation with mutant RNAPII genes. Upon transformation with the mutant *rpb* constructs the cells were selected on G418 plates and a high proportion of the G418 resistant transformants display the same defects seen in the original mutant strains.

For all mutants tested, a proportion of the recovered transformants displayed phenotypes similar to the original strain. Two of the regenerated mutant colonies that exhibited silencing, and ts phenotypes if applicable, for each

of the mutants were sequenced and it was confirmed that the same mutations were present. Comparative plating assays for silencing and temperature sensitivity confirm that these regenerated mutants have the same silencing defects and temperature sensitivity as the original mutants (Figure 3.3 and Figure 3.4).

In general, the newly isolated *rpb3* and *rpb11* mutants show strong defects in silencing of the *ade6*$^{+}$ marker gene located in pericentromeric heterochromatin. The random mutagenesis strategy employed was therefore successful in generating new RNAPII mutants with effects on heterochromatin integrity. Mutations in 10 unique residues have been identified and confirmed to cause silencing defects in these mutants. Three new temperature sensitive alleles of *rpb3* were isolated (*rpb3-1A, rpb3-1B/20B, rpb3-8E*), two of which have very strong silencing defects at the permissive temperature.

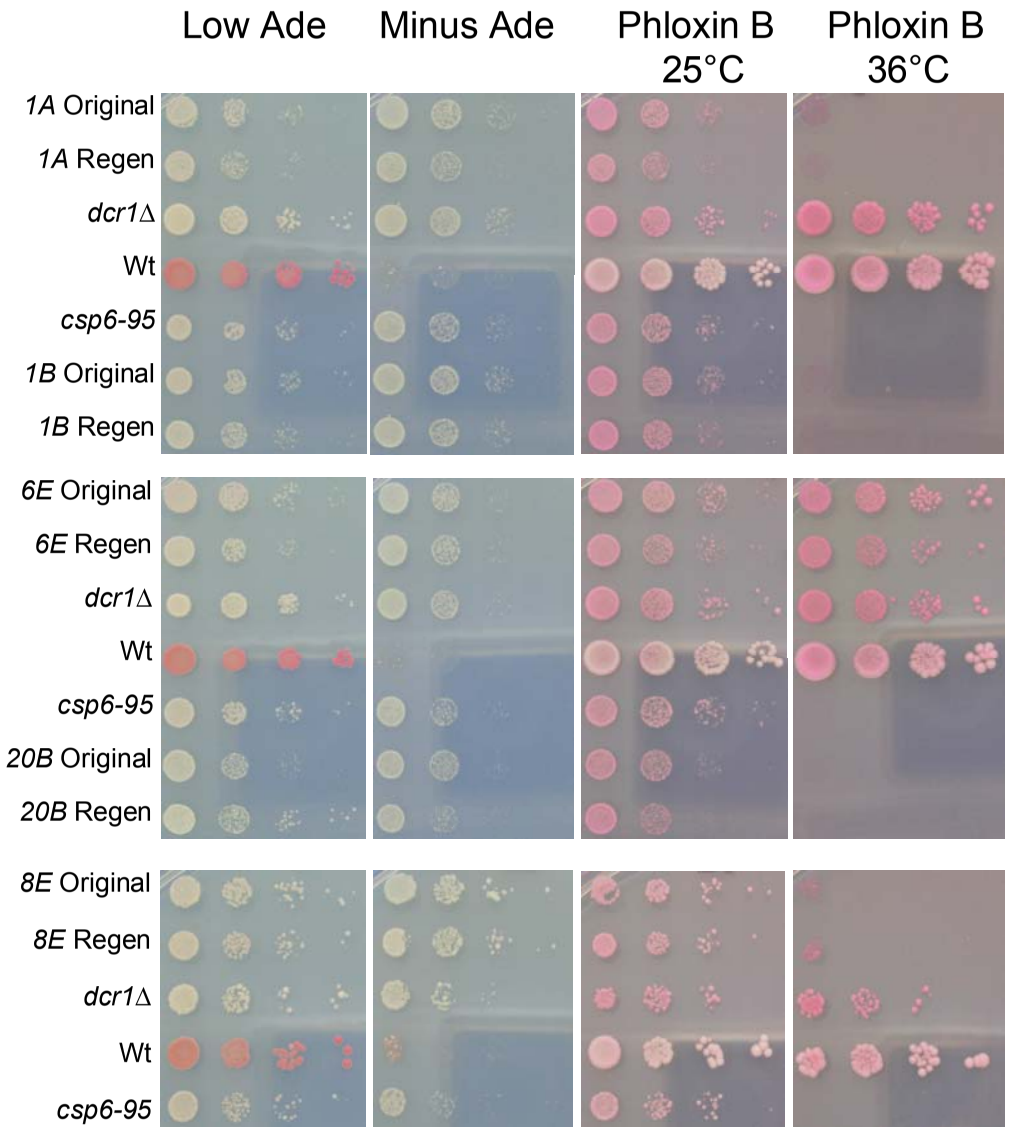
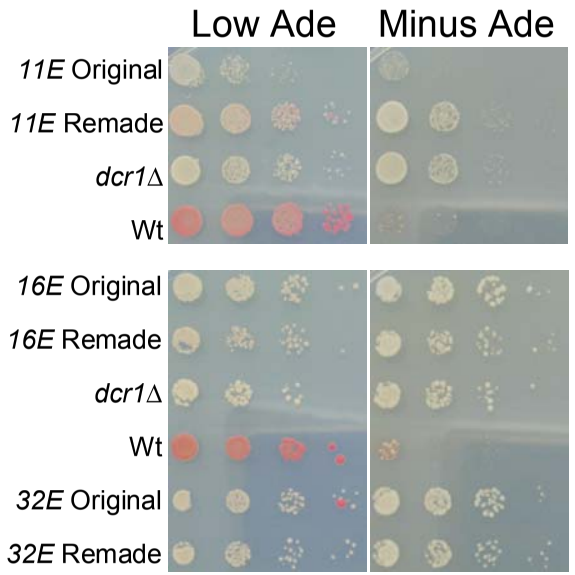


Figure 3.3 – The regenerated *rpb3* mutants display the same phenotypes as the original mutant strains. Comparative plating assays display the regenerated phenotypes of the *rpb3* mutants. The *ade6+* gene is positioned in the pericentromeric outer repeats where the gene is normally repressed by heterochromatin. Growth on low adenine plates provides colour readout of *ade6+* gene expression where white colonies have *ade6+* expressed (defective silencing) and the red colonies have wildtype repression. On minus adenine plates only cells expressing *ade6+* can survive. Phloxin B is a red pigment which is only accumulated in dead cells as live cells actively pump the pigment out of the cells.



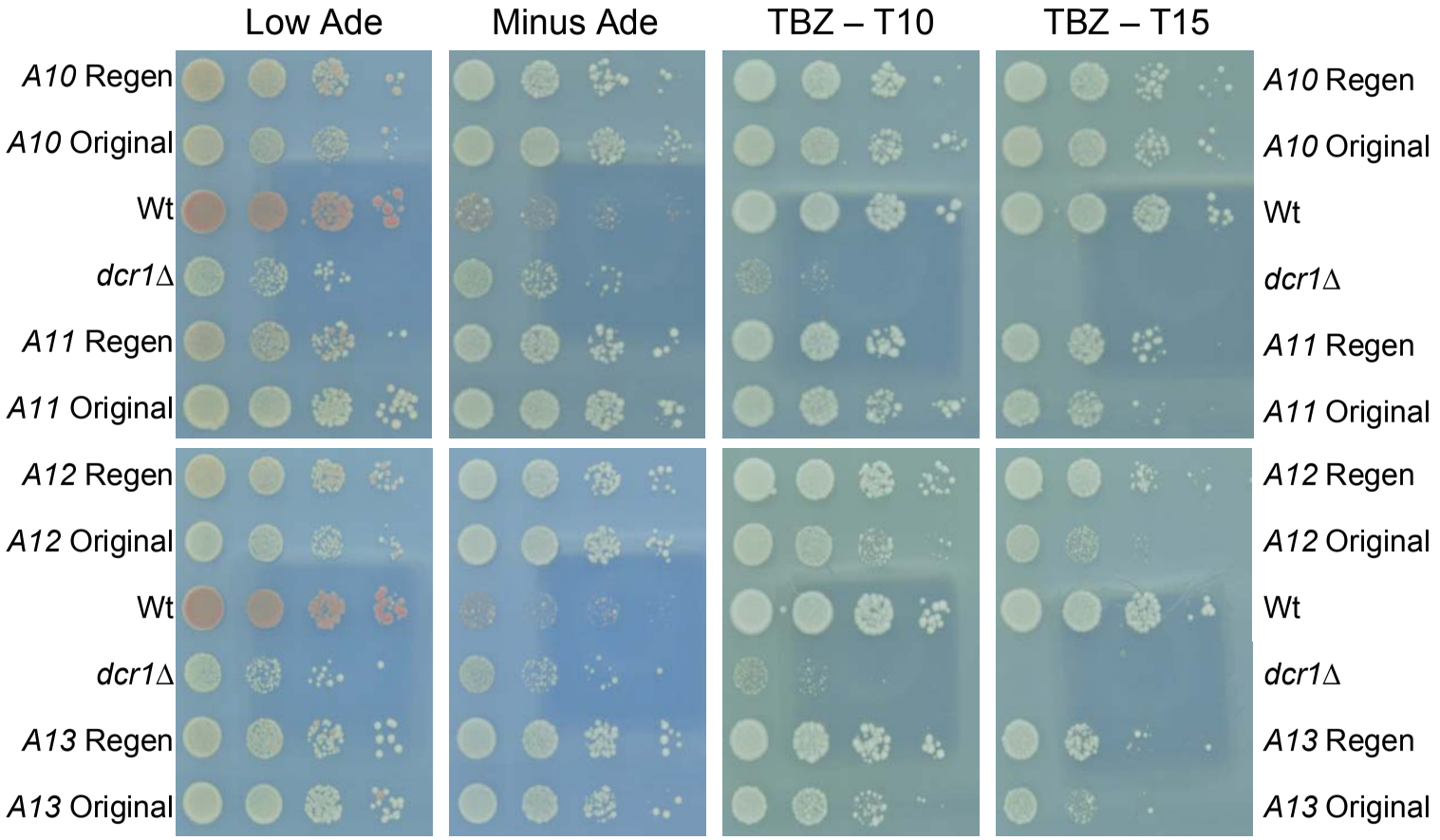


Figure 3.4 – The regenerated *rpb11* mutants display the same phenotypes as the original mutant strains. Comparative plating assays display the regenerated phenotypes of the *rpb11* mutants. Expression of the *ade6+* gene causes colonies to have a white colour whereas repression of *ade6+* causes a red pigment to accumulate. The *ade6+* gene is located in the pericentromeric heterochromatin thus expression of the gene indicates defective silencing in the region. Growth on plates with no adenine is indicative of defective silencing. A defect in pericentromeric heterochromatin results in lagging chromosomes and sensitivity to the microtubule destabilising drug TBZ.

3.2.5 Positions of the affected residues in *rpb3* and *rpb11* mutations relative to protein structure

The 3° structure of the RNAPII complex is highly characterised in *S. cerevisiae*, including crystal structures of the complex at a resolution of 3.8Å. By mapping the mutations for my RNAPII mutants on the reference structure I hoped to gain insight into the molecular basis of the observed defects.

The α-motifs are regions of the prokaryotic RNA Polymerase α subunit that are conserved in the eukaryotic polymerases (Ulmasov et al., 1996). The α-motif is highly conserved from prokaryotes to humans and can be seen in both the eukaryotic RNAPII α subunit homologs: Rpb3 and Rpb11 (Figure 3.5). The *rpb11-A13* and *rpb11-A10* mutations are both found in the α-motif 1. Benga et al 2005 previously made three mutations in the start of the α-motif of RPB11 in *S. cerevisiae,* substituting E38, D39 and L42 for alanine residues, both individually and in combination. Equivalent mutations were also made in the human homologs: hRpb11-a and hRpb11-b1 (Benga et al., 2005). As Rpb11 normally forms a heterodimer with Rpb3 in assembled RNAPII, they tested the newly made Rpb11 mutant proteins for their ability to heterodimerise with Rpb3 using the yeast two-hybrid (Y2H) technique. They found that the yeast subunits were unaffected for forming heterodimers, but the human subunits had reduced ability to make heterodimers. As the *rpb11-A13* mutation is in the same region as the mutated residues it is possible that it could affect heterodimerisation, although the fact that - similar to the budding yeast mutants seen by Benga et al 2005 - the mutant is healthier than the other RNAPII mutants suggests dimerisation is unlikely to be severely affected. In the case of the *rpb11-A10* mutation (K61E) it may be affecting interactions with Rpb1 due to its location (Figure 3.6A) and its change from a positively charged amino acid to a negatively charged one.

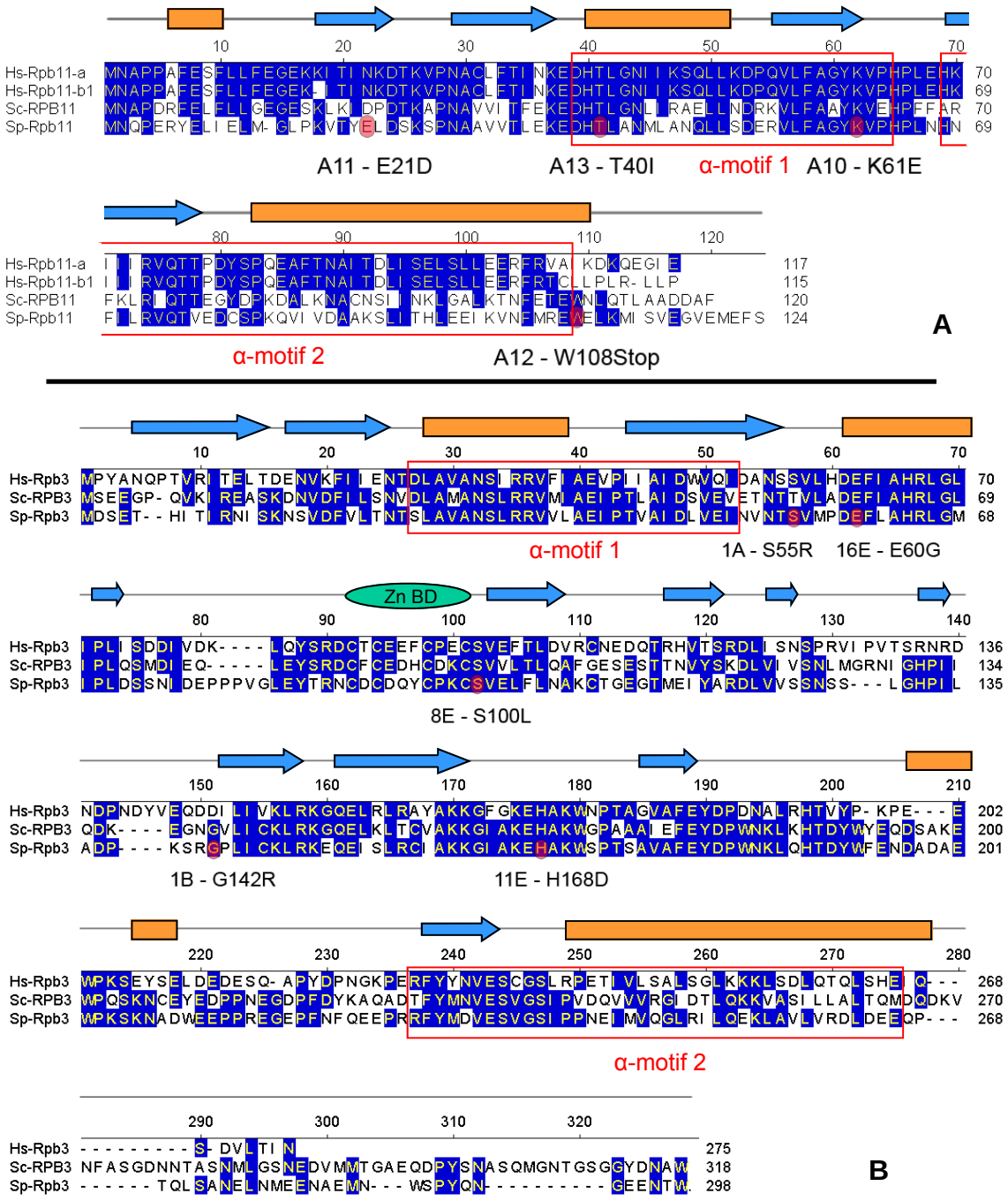


Figure 3.5 – Sequence alignments of the RNAPII subunits Rpb3 and Rpb11 from *Homo sapiens* (Hs), *Saccharomyces cerevisiae* (Sc) and *Schizosaccharomyces pombe* (Sp). The 2° structure is depicted above the sequences with β-sheets as blue arrows and α-helices as orange boxes. Residues highlighted in pink mutated in the new RNAPII mutants. The highly conserved α-motifs which bear homology to the bacterial α subunit are highlighted by red boxes (Ulmasov 1996). (**A**) Alignment of the Rpb11 subunits. There are two homologues in humans compared to the single homologues seen in yeast. (**B**) Alignment of the Rpb3 subunits. The Zinc binding domain (Zn BD) in Rpb3 is one of seven Zn BD sites in eukaryotic RNAPII (Donaldson 2000).

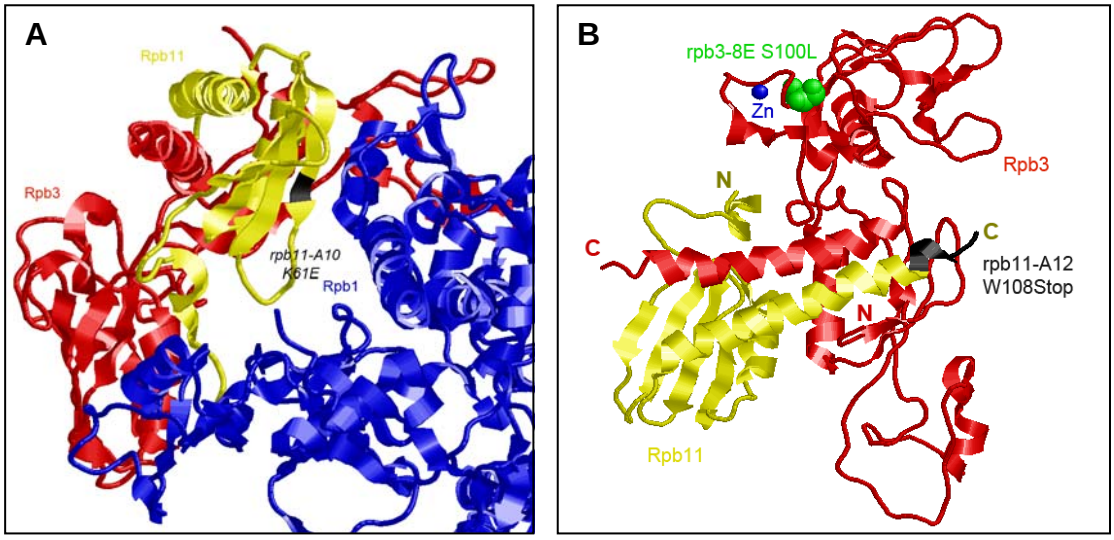


Figure 3.6 – The protein position of the new RNAPII mutations. (**A**) The *rpb11-A10* mutation is found on a β-sheet which interacts with an α-helix from Rpb1. (**B**) The Rpb3-Rpb11 dimer. The truncated version of Rpb11, due to the nonsense mutation of *rpb11-A12*, is depicted with the lost residues highlighted in black. The S100 residue which is mutated in *rpb3-8E* is shown in green spacefill adjacent to the Zinc Binding Domain.

Benga et al 2005 also investigated heterodimerisation of Rpb3 and Rpb11 by Y2H using truncated versions of Rpb11. The *rpb11-A12* mutation (W108Stop) causes a truncation of Rpb11 (Figure 3.6B) similar to the truncation mutants 1-106 made for the budding yeast and human proteins. Benga et al 2005 found that there was a significant reduction in the ability to dimerise in the truncation proteins that lost a portion of the C-terminal α-helix and loss of the entire C-terminal α-helix resulted in complete loss of dimerisation. Since the *rpb11-A12* mutation at W108 is very near the end of the C-terminal α-helix, and therefore it is mainly the unconserved C-terminal region (Figure 3.5) that is lost, these analyses suggest that *rpb11-A12* may have a slight reduction in the ability to create Rpb3-Rpb11 dimers, but is unlikely to completely lose dimerisation. Dimerisation of Rpb3 with the *rpb11-A12* truncated Rpb11 protein could be tested in the same way using Y2H.

Rpb3-1A and *rpb3-1B* have the strongest silencing defects of all the new RNAPII mutants as well as being ts and the sickest of the mutants. Although the

mutated residues are separated by 85 amino acids in the 1° structure, they are in fact quite close in the 3° structure with both residues on the surface that interacts with Rpb10, suggesting that this interaction might be affected (Figure 3.7A).

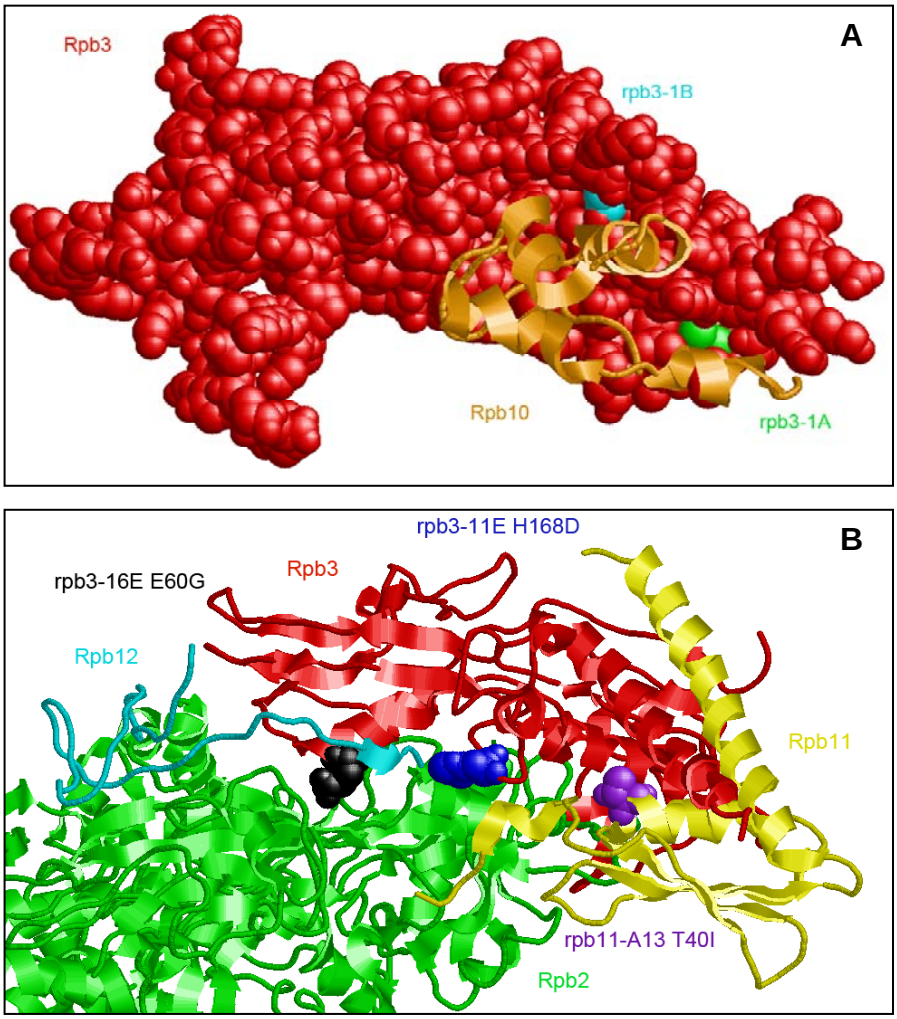


Figure 3.7 – The protein position of the new RNAPII mutations. (**A**) The sickest mutants, *rpb3-1A* and *rpb3-1B*, have their mutations situated close to the Rpb3-Rpb10 interface. (**B**) The *rpb3-11E* and *rpb3-16E* mutations are both on the interface with Rpb2 and Rpb12. The *rpb11-A13* mutation is situated in an α-helix which runs parallel to an α-helix of Rpb3.

The *rpb3-8E* mutation S100L is adjacent to the Zinc binding domain (Zn BD) (Donaldson and Friesen, 2000) which is made up of four co-ordinating cysteine residues (C90, C92, C96, C99) (Figure 3.6B). As this mutation is the

substitution of a polar residue for a nonpolar one it is likely that the 3° structure is affected, which may well disrupt the Zn BD. It is known that there are seven Zn BD sites in the RNAPII complex and that Zinc is essential for RNAPII function (Donaldson and Friesen, 2000). Disruption of the Rpb3 Zn BD may well therefore account for the phenotype of the *rpb3-8E* mutant. In order to test this theory one or more of the cysteine residues that make up the Zn BD could be mutated and the resulting mutants tested, alongside *rpb3-8E* mutant, for ts and silencing defects. A Zn binding assay (Schiff et al., 1988) could also be used to test if Zn binding is disrupted in the *rpb3-8E* mutant. The Zn BD of Rpb3 provides part of an interaction surface which has been shown to be important in budding yeast for non-basal transcriptional activation (Tan et al., 2000) and to provide a contact site for the Mediator subunit Med7 (Soutourina et al., 2011). In *S. cerevisiae* the cysteine substitution mutant C92R does not cause a ts phenotype but does have a synthetic ts phenotype with the A159G mutation which is situated on a surface beside the Zn BD (Tan et al., 2000). This suggests that the S100L mutant in *S. pombe* causes more disruption to the interaction surface or that the Zn BD in *S. pombe* has a more important function than in *S. cerevisiae*.

The *rpb3-11E* and *rpb3-16E* mutations are both in highly conserved amino acids (Figure 3.5) which are situated on the interface of Rpb3 with Rpb12 and Rpb2 (Figure 3.7B). The *rpb3-16E* mutation site, which is incidentally also the residue mutated in *rpb3-6E* and *rpb3-32E*, is effectively closed from the surface of the RNAPII complex by Rpb12 – the *rpb3-16E* mutation may well therefore disrupt the interaction with this subunit. Integrity of the Rpb3-Rpb12 interaction could be tested using a Y2H experiment comparing interaction of wildtype and mutant Rpb3 proteins with Rpb12. The *rpb3-11E* mutation, H168D, is the substitution of a basic amino acid for an acidic residue. The residue H168 is located in a pocket of predominantly positively charged residues (Figure 3.8). The substitution of charges in *rpb3-11E* may have an affect on binding/interaction on this surface of the RNAPII complex. This theory is difficult to test biochemically however because the surface is formed by a number of RNAPII subunits and there are no obvious candidates for binding the surface.

Mutation of the basic residues in the Rpb2 subunit on this surface could provide some insight into whether these residues have a cumulative effect on silencing with the *rpb3-11E* mutation. Alternatively mass spectrometry could be used to compare the binding partners of Rpb3 in wildtype and mutant backgrounds in order to find candidates for binding of the surface.

The positions of the new mutations in the RNAPII complex hint towards the residues being involved in certain functions such as subunit-subunit binding or metal binding. Most of the mutations are found at sites facing inwards towards the RNAPII complex rather than on the exterior surfaces. The exceptions to this are the nonsense mutation of *rpb11-A12* which causes the loss of the unstructured C-Terminal region of Rpb11 and, to a lesser extent, the *rpb3-11E* mutation which is partially open to the outer surface in a positively charged region (Figure 3.8).

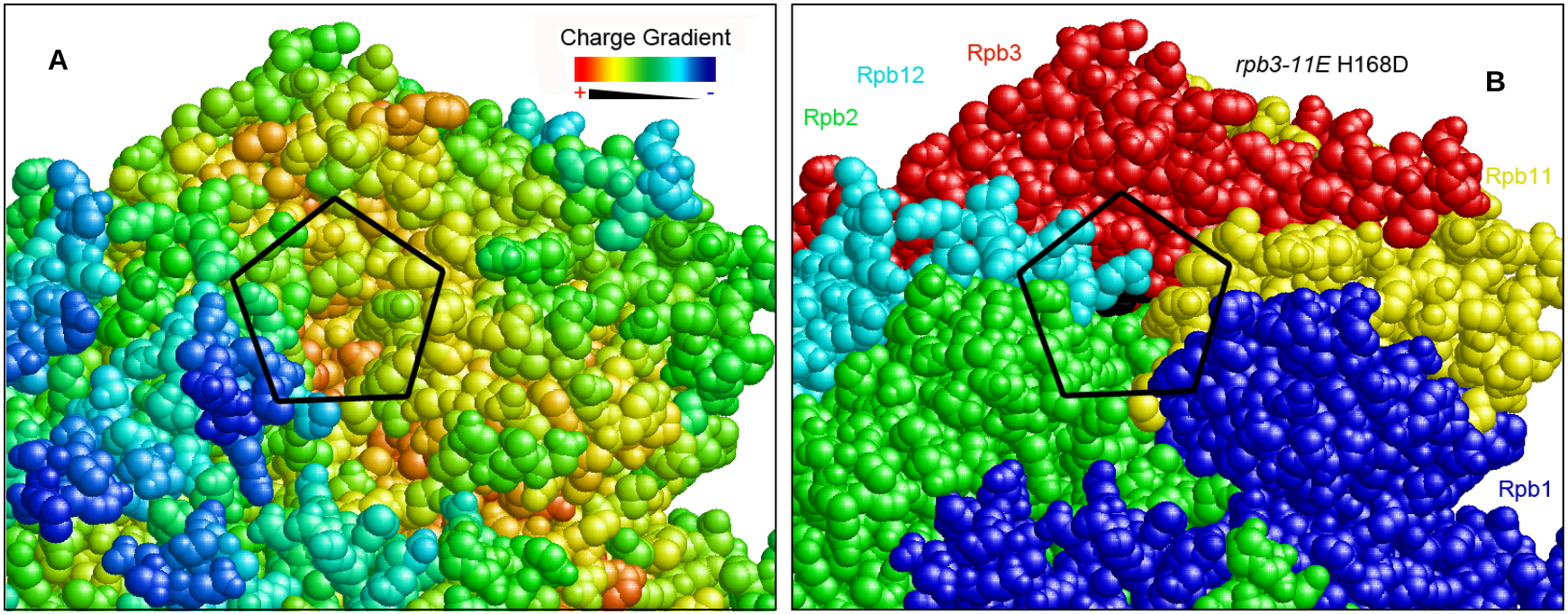


Figure 3.8 – The protein position of the mutated amino acid in *rpb3-11E*, H168D. (**A**) RNAPII residues coloured according to charge. H168 is a positively charged (orange) residue in a predominately positively charged region of the complex. The H168D mutation substitutes a positive charge for a negative charge. (**B**) RNAPII coloured according to subunits with the *rpb3-11E* mutation (H168D) coloured black. H168 is situated at the interface of Rpb3 with Rpb12 and Rpb2.

3.3 Discussion

In this chapter I have attempted to generate new RNA Polymerase II (RNAPII) mutants. Initially I took the approach of recreating known mutations by targeted mutagenesis, however this strategy did not result in isolation of any mutants with a temperature sensitive (ts) phenotype. Subsequently I employed a random mutagenesis approach, which was successful in producing three new ts alleles of the *rpb3* gene and a number of mutants of *rpb3* and *rpb11* with pericentromeric heterochromatin defects. The 12 mutants with the strongest phenotypes were sequenced and the mutations seen in the RNAPII subunit gene were confirmed to cause the defects. The mutations were mapped to the protein structure in the context of the RNAPII complex in order to gain possible insight into the mechanism of the defects.

The random mutagenic screens for Rpb3 and Rpb11 were successful in generating new RNAPII mutants, with a combined total of 352 potential mutants. Four *rpb3* and two *rpb11* ts mutants were produced, but unfortunately none of these was ideal for use as conditional switches for RNAPII transcription due to their slow growth at the permissive temperature, which indicates that the mutations affect transcription even at 25°C. The Rpb3 mutagenesis screen was on a relatively small scale compared to the Rpb11 screen, however, despite only finding 35 *rpb3* mutants, three mutants were found to harbor mutations of the same amino acid, and two of the ts mutants were found to have the same nucleotide mutation. This redundancy suggests that the screen achieved saturation.

Another strategy for selecting RNAPII ts mutants would be random mutagenesis of the two largest subunits, *rpb1* and *rpb2*. The size of these genes means that it would not be feasible to cover an entire gene through random mutagenesis in the same way as for *rpb3* and *rpb11*. Unfortunately the repetitive nature of the *rpb1* C-terminal Domain (CTD) makes the 3' end of *rpb1* difficult to manipulate. However the closest gene to the 5' end of *rpb1* is >700bp upstream so there would be space to insert the marker gene at the 5' end and still avoid disturbing the *rpb1* promoter, thus random mutagenesis could be carried out for

the N-terminal of *rpb1*. The *rpb2*$^{+}$ gene is closely flanked on either side by its neighbouring genes however the 3' end could allow insertion of the marker gene because the genes are convergent and thus only the terminator regions would be disrupted, one of which can be replaced by the terminator on the marker construct whilst the other is left intact.

The fact that most of the RNAPII mutants that were sequenced had mutations in residues internal to the RNAPII complex along with the generally reduced growth rates of the mutants suggests that the mechanism of silencing defect seen is likely due to defective interaction of the RNAPII subunits. Despite this the mutants may still be useful in dissecting the role(s) of RNAPII in the RNAi pathway; for example if binding of a particular subunit is partially disrupted in a mutant then the subunit is implicated as important for the RNAi pathway, perhaps as an interaction surface.

Characterisation of the silencing defect phenotypes is necessary to determine at which step in the RNAi pathway the mutants affect. Of particular interest is whether the mutants have defects in the transcription of the pericentromeric outer repeats which are required to generate siRNA (like *rpb7-G150D*) or whether they have a non-transcriptional role in RNAi (like *rpb2-m203*). Currently the only evidence for a non-transcriptional role of RNAPII in RNAi is through the disruption of the RNAi pathway in *rpb2-m203* (Kato, 2005). As the processing of transcripts into siRNAs occur in-cis it is possible that the RNAPII complex is acting as a platform for RNAi, similar to its role in splicing. Alternatively RNAPII may affect the RNAi pathway in a more indirect way through its interactions with intermediate proteins such as chromatin remodelling enzymes. The characterisation of the new RNAPII mutants is described in the next chapter.

Chapter 4 – Characterising RNAPII mutants

4.1 Introduction

The RNAPII mutants that were generated and identified in Chapter 3 show a spectrum of silencing and growth defects. This range of phenotypes suggests that defects in different mutants are caused either by different mechanisms, or a common mechanism with a range of severity. In order to distinguish between these possibilities it is important to learn more about the silencing defects. Currently there are two known RNAPII mutants that affect the integrity of pericentromeric heterochromatin, neither of which affects the integrity of heterochromatin at any other regions (eg mating-type locus or telomeres). This is significant because RNAi is essential for heterochromatin formation at the pericentromere whereas the other heterochromatic regions have redundant pathways for maintaining heterochromatin in addition to RNAi.

rpb7-G150D (aka *csp3*) is a slow growing temperature sensitive (ts) mutant which has a variegating silencing phenotype due to defective heterochromatin (Ekwall, 1999). Its silencing defect is associated with reduced levels of transcription of the centromeric non-coding RNAs, which are normally diced into siRNAs (Djupedal, 2005). The resulting low levels of siRNAs are believed to cause inefficient RNAi and subsequently the loss of silencing at the pericentromeric outer repeats. Mutant cells are ts and grow more slowly than wildtype cells because global transcription is affected, presumably affecting essential pathways as well as the non-essential RNAi pathway.

rpb2-m203 grows relatively normally but displays variegating silencing indicative of unstable heterochromatin (Kato, 2005). It exhibits reduced siRNA levels with a corresponding accumulation of centromeric repeat transcripts. Together these phenotypes suggest that RNAPII has an additional role in silencing the pericentromere repeats downstream of transcription. Thus far there has been no further analysis of how RNAPII influences the formation of centromeric heterochromatin.

Based on the phenotypes of the two published RNAPII mutants that have silencing defects, it would be expected that the new RNAPII mutants will fall into one of three categories:

(1) As with *rpb7-G150D*, such mutants will have a defect in the production of transcripts from the centromeric outer repeats and other genes. Such mutants are likely to exhibit slow/reduced growth due to general transcription defects.

(2) As in *rpb2-m203*, such mutants will transcribe centromere repeats normally but the transcripts will not be processed into siRNA causing centromere transcript levels to accumulate. Such mutants will affect heterochromatin integrity and exhibit reduced H3K9 methylation levels specifically at the pericentromere outer repeats.

(3) Mutants that do not fit into either of the above categories.

In order to assess the silencing defects for the new mutants they have been tested for levels of centromeric repeat transcript, siRNA, H3K9me2 and silencing at the mating-type locus and the telomeres.

4.2 Results

4.2.1 RNAPII mutants display variable levels of pericentromeric transcripts and siRNAs

The two previously characterised RNAPII mutants have had differing phenotypes in terms of transcription of the centromeric repeats and processing into siRNAs therefore the newly generated RNAPII mutants were first investigated for defects in transcription of the repeats and siRNA processing. Centromeric transcripts are normally converted to dsRNA which is cleaved into siRNAs therefore it is only possible to determine the steady state level of these transcripts in cells defective in RNAi (eg *dcr1*Δ which lacks the RNase Dicer). Quantitative Reverse Transcription (qRT)-PCR was used to determine centromere transcript abundance in the RNAPII mutants, whilst siRNA levels were assessed by northern analyses (Figure 4.1), using RNA prepared from the

same cell lysates. The mutants that were tested were chosen based on their phenotypes: (1) *rpb3-1A* and *rpb3-1B* have the strongest silencing defects but are likely to be affecting essential pathways as they are very slow growing and ts. (2) *rpb3-11E, rpb3-16E, rpb11-A10, rpb11-A11, rpb11-A12* and *rpb11-A13* have quite strong silencing defects and are less likely to be affecting general transcription as they display relatively good growth. (3) *rpb3-8E* is unlike the other mutants as although it has a weaker silencing defect, it grows more readily than the other ts mutants and this suggests that it might have a different mechanism of defect compared to the other mutants.

The mutants present a range of transcript accumulation levels and siRNA levels. The results of Figure 4.1 are summarised and compared in Table 4.1 in order to assess whether the phenotypes observed are due to defective transcription or their processing to siRNA. In two mutants (*rpb3-1A* and *rpb11-A12*) low transcript levels were detected and siRNA levels were similar to wildtype cells. Since transcript and siRNA abundance are unaffected this suggests that the phenotypes seen in *rpb3-1A* and *rpb11-A12* are downstream of transcription and siRNA generation and thus likely to involve a chromatin modification defect. *rpb11-A13* appears to resemble *rpb2-m203*, with increased transcript accumulation and decreased levels of siRNAs, while *rpb11-A11* has low levels of both transcripts and siRNAs, similar to *rpb7-G150D*. The other mutants have intermediate phenotypes suggestive of partial defects in either transcription or processing to siRNA.

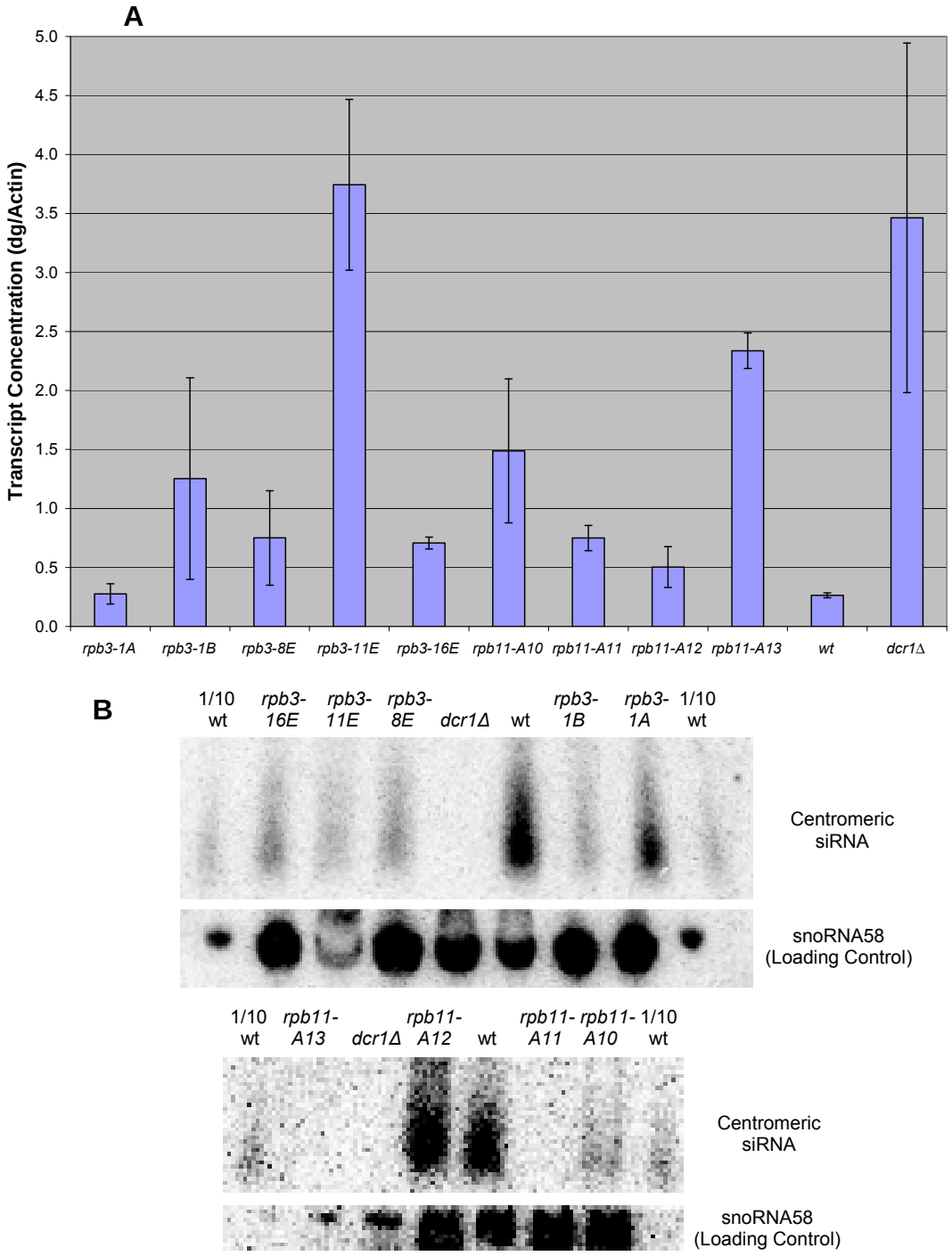


Figure 4.1 – RNAPII mutants display a range of pericentromeric transcript and siRNA processing defects. The long RNA and siRNA samples were extracted from the same culture using different precipitation conditions to separate by size. The RNAi mutant *dcr1*Δ causes loss of processing of long RNA into siRNA resulting in accumulation of cen transcripts compared to wild-type and complete loss of siRNA. (**A**) Semi-quantitative qRT-PCR analysis of reverse transcribed centromeric (dg) transcripts. A single RNA sample for each strain was reverse transcribed in three separate reactions and each reaction was quantified in triplicate by Real Time PCR to produce the observed error bars. (**B**) Centromeric siRNA levels in the RNAPII mutants detected by northern analysis.

Strain	Cen Transcript	Cen siRNA	Overall RNA Phenotype
Wildtype	Low	High	Wildtype
***dcr1*Δ**	High	Low	Defective RNA Processing
1/10th Wt	N/A	Intermediate	
rpb3-1A	Low	High	Wildtype
rpb3-1B	Low/Intermediate	Low/Intermediate	Partially Defective Transcription
rpb3-8E	Low	Intermediate	Partially Defective Transcription
rpb3-11E	High	Low/Intermediate	Partially Defective RNA Processing
rpb3-16E	Low	Intermediate	Partially Defective Transcription
rpb11-A10	Intermediate	Intermediate	Partially Defective RNA Processing
rpb11-A11	Low	Low	Defective Transcription
rpb11-A12	Low	High	Wildtype
rpb11-A13	High	Low	Defective RNA Processing

Table 4.1 – Comparison of pericentromeric transcript and siRNA levels using the results displayed in Figure 4.1. The results displayed in Figure 4.1 were used to produce 'High', 'Intermiediate' or 'Low' scores for the levels of centromeric transcripts and siRNAs summarised in this table. The long RNA and siRNA samples were extracted from the same culture using different precipitation conditions to separate by size. The RNAi mutant *dcr1*Δ causes loss of processing of long RNA into siRNA resulting in accumulation of cen transcripts compared to wild-type and complete loss of siRNA. The centromeric (dg) transcripts were measured by semi-quantitative qRT-PCR analysis of reverse transcribed RNA. A single RNA sample for each strain was reverse transcribed in three separate reactions and each reaction was quantified in triplicate by Real Time PCR. The centromeric siRNA levels in the RNAPII mutants were detected by northern analysis.

However, upon repeating northern analyses of siRNA different results were obtained which are summarised in Table 4.2. As the *ade6*$^{+}$ marker silencing observed on low adenine plates for the RNAPII mutants shows variegation, one explanation is that the siRNA variability may reflect phenotypic variegation – different degrees of silencing would explain differences in siRNA levels observed in distinct RNA preparations. To test this possibility a selection assay was used. Four of the RNAPII mutants were chosen that were most likely to have silencing defects not linked to transcription, based on their growth rates and the phenotypes seen in Table 4.2: *rpb3-1A, rpb3-11E, rpb11-A10* and *rpb11-A13*. The *rpb2-m203* mutant also displays variegation of silencing; for this reason *rpb2-m203* was also chosen for comparison of variegation with the new RNAPII mutants.

The RNAPII mutants were crossed into a background in which the cen1:*ade6*$^{+}$ gene inserted in the outer repeats of cen1 was replaced by the *ura4*$^{+}$ gene. The advantage of *ura4*$^{+}$ is that it allows selection and counter-selection for cells expressing or not expressing *ura4*$^{+}$. Thus populations of cells can be selected when *ura4*$^{+}$ is expressed or silenced by growth in media lacking uracil, or supplemented with counter-selective 5-fluoroorotic acid (5-FOA), respectively. In several mutants the FOA^{R} cell populations exhibited increased siRNA levels (relative to populations grown in –URA), consistent with selective pressure for *ura4*$^{+}$ silencing. However, the most extreme variation in siRNA levels seen previously for *rpb3-1A* (Table 4.2) was not reflected in such selection experiments (Figure 4.2, Table 4.3).

Experiment	1	2	3	4
Strain				
Wildtype	+++++	+++++	+++++	+++++
*dcr1*Δ	-	-	-	-
rpb2-m203	N/A	N/A	N/A	++++
1/10th Wt	N/A	++	++	++
rpb3-1A	++++	-	++++	+
rpb3-1B	++	N/A	++	++++
rpb3-8E	N/A	+	+++	+++
rpb3-11E	N/A	+++	+++	+
rpb3-16E	N/A	++++	+++	++
rpb11-A10	N/A	N/A	++	+++++
rpb11-A11	N/A	+	-	++++
rpb11-A12	-	N/A	+++++	+
rpb11-A13	-	-	N/A	-

Table 4.2 – Summary of northern analyses for siRNA levels. Four sets of RNA samples were prepared on four different occasions. The RNAi mutant *dcr1*Δ causes loss of processing of the long pericentromeric transcripts into siRNA resulting in complete loss of siRNA. The siRNA levels appear to be inconsisitent for the majority of the RNAPII mutants tested. The levels of siRNA have been given an arbitrary score based on the levels observed by northern analyis.

Strain	Wt	*dcr1*Δ	*clr4*Δ	1/10th Wt	*rpb3 -1A*	*rpb3 -11E*	*rpb11 -A10*	*rpb11 -A13*	*rpb2 -m203*
Media									
-URA	+++++	-	+	N/A	-	+	+	++++	-
5-FOA	++++++	N/A	N/A	++	+	+++	++	++++++	-

Table 4.3 – Summary of the siRNA levels seen in the selection assay by northern analysis (Figure 4.2). This table allows comparison to the variable results of the previous northern blots for siRNA levels summarised in Table 4.2. The levels of siRNA have been given an arbitrary score based on the levels observed by northern analyis in Figure 4.2. The RNAi mutant *dcr1*Δ causes loss of processing of the long pericentromeric transcripts into siRNA resulting in complete loss of siRNA. The HMT mutant *clr4Δ* causes loss of methylation of H3K9 resulting in global loss of heterochromatin and reduced levels of siRNA.

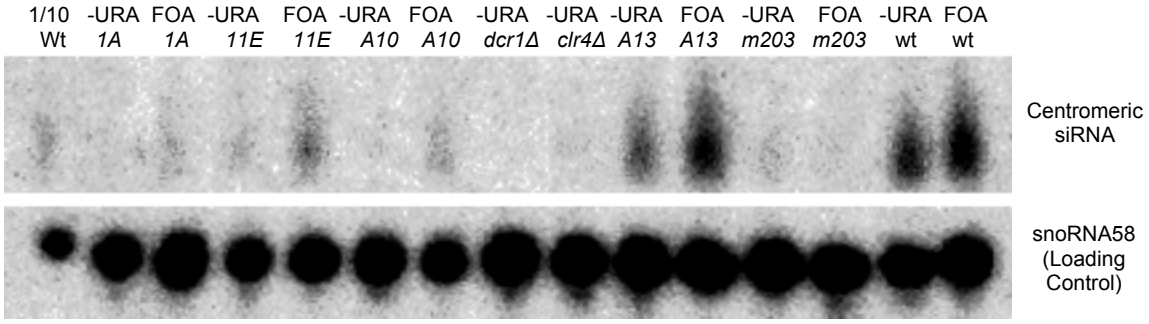


Figure 4.2 – siRNA levels are affected by selection of *cen1:ura4* marker gene expression. Selection assay testing variegation of silencing using northern analysis to determine centromeric siRNA levels in the RNAPII mutants when grown in -URA (selecting for *ura4*+ expression) and 5-FOA (selecting against *ura4*+ expression)

In fact the *rpb11-A13* mutants bearing cen1:*ura4*$^{+}$ at the outer repeats has levels of siRNA that are within range of the wildtype levels when grown in either 5-FOA or –URA media, compared to the complete loss of siRNA originally detected on several analyses. The fact that selection, for silencing and non-silencing of the marker gene at the heterochromatic outer repeats of the centromere, does not result in both the previously observed highest and lowest levels of siRNA for each mutant is unexpected. This result suggests that there is another aspect apart from short term heterochromatin assembly/disassembly

that is affecting the variability of siRNAs in these mutants. In order to check if different populations of cells for a particular mutant have variable levels of siRNA, numerous isolates of *rpb11-A13* were tested simultaneously (Figure 4.3).

The analysis of these different *rpb11-A13* strains indicates that although the original strain of *rpb11-A13* does have low levels of siRNA as originally shown, the levels vary in different isolates. Both of the *rpb11-A13* strains with cen1:*ade6*$^{+}$ were made by transformation of the tester strain the mutated *rpb11* construct DNA, whilst the strains with cen1:*ura4*$^{+}$ were made by crossing the original *rpb11-A13* strain. It is possible that the progression through meiosis effectively resets the heterochromatic state of the pericentromeric repeats or changes the balance of RNAi in some manner. Another explanation for the differences in strains may be due to the number of divisions through which the strains have been passaged. It has been observed in the lab that different strains of the same RNAi deletion mutant can have differing levels of siRNA (personal communication A. Buscaino). It has been suggested that this may be due to the number of divisions since the deletion was created whereby the deletion mutant initially has partially defective heterochromatin which is gradually lost with each division. The original *rpb11-A13* strain was grown on plates for longer after creation before being frozen for storage compared to the newer strains, which would be consistent with gradual loss of silencing over time although this was not directly tested.

Mutations which affect pericentromeric heterochromatin integrity often have varying degrees of loss of H3K9 methylation compared to siRNA levels, depending on which point in the RNAi-dependent pathway they are involved. For example *dcr1Δ* (deletion of the RNase responsible for siRNA generation), which loses all siRNAs, can retain varying levels of H3K9 methylation levels (5-30%) on the outer repeats, whereas cells lacking the H3K9 methyltransferase Clr4 (*clr4*Δ) are devoid of all H3K9 methylation at all loci and have barely detectable levels of siRNAs. In order to see where the RNAPII mutants fit into the pericentromeric heterochromatin formation pathway it is important to know the effects of the mutants on H3K9 methylation.

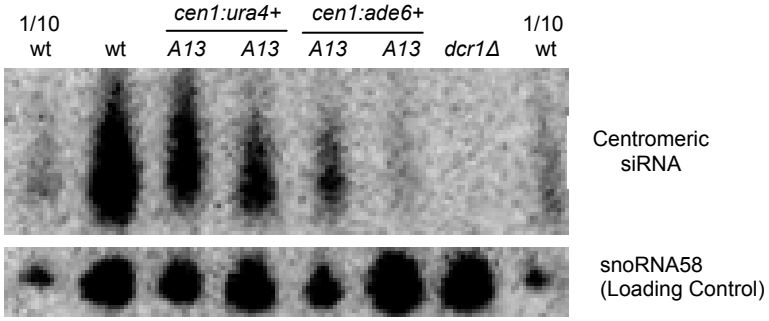


Figure 4.3 – siRNA levels in the *rpb11-A13* mutant vary widely in different isolates. Northern analysis detecting centromeric siRNA levels in numerous strains of *rpb11-A13.* Four different *rpb11-A13* strains are tested here, two strains containing the *ura4+* marker gene inserted in the pericentromeric outer repeats and two strains containing the *ade6+* marker gene at the outer repeats.

4.2.2 Heterochromatin integrity in RNAPII mutants

To determine if heterochromatin remains intact in the RNAPII mutants, one of the hallmarks of heterochromatin, H3K9 methylation, was checked by Chromatin Immunoprecipitation (ChIP) (Figure 4.4). In order to provide comprehensive analyses of mutant phenotypes, cells for ChIP analysis were taken from the same cultures as those used for RNA analysis (experiment 4 Table 4.2). The previously characterised RNAPII mutant, *rpb2-m203*, was tested simultaneously for comparative purposes.

The H3K9me2 levels for all of the RNAPII mutants are at least as high as the wildtype strain, suggesting that the centromeric heterochromatin integrity remains intact in these mutants. Although there are apparently higher levels of H3K9me2 in most of the mutants compared to wildtype, this seems counter-intuitive when compared with the RNAi defects observed (Figure 4.1, Chapter 3 Table 3.1 and Table 3.2). The phenotypes observed for *rpb3-11E* are a good example of this mismatch of phenotypes: high accumulation of repeat transcript and low siRNA levels reminiscent of an RNAi mutant but high levels of H3K9me2 which would suggest fully intact heterochromatin or in fact an enhanced silencing phenotype. *rpb2-m203* was found to have low levels of H3K9me2 on the outer repeats, consistent with the previously published data

(Kato, 2005). The *rpb2-m203* H3K9me2 levels seem to fluctuate which is a similar phenotype to RNAi mutants like *dcr1Δ*.

The H3K9me2 levels were also assessed (Figure 4.5) in the mutant cells that were grown in 5-FOA and –URA for the variegation assay previously described in Figure 4.2. A clear correlation is seen between selective and counter-selective media and the levels of H3K9me2. Higher levels of H3K9me2 were detected on the pericentromeric dg repeats when cells are grown in -URA than 5-FOA. Surprisingly, the correlation is opposite to what would be expected, since the absence of supplementing uracil should provide selective pressure for increased *ura4*$^{+}$ expression (and thus loss of heterochromatin). A possible explanation for this is that the pressure for alleviating the silencing of the *ura4*$^{+}$ gene in the pericentromeric repeats is causing RNAi/ClrC factors to be released from the marker gene which are then free to intensify silencing of the adjacent pericentromeric repeats. However this explanation does not fit well with the corresponding siRNA data (Figure 4.2), which shows a clear decrease in siRNA in the strains grown in –URA compared to 5-FOA, indicating decreased silencing on the repeats. The disparity between siRNA levels and the consequent H3K9 methylation levels on the pericentromeric repeats in this selection assay is also consistent with the results seen for the assays done in non-selective media (Figures 4.1 and 4.4).

Another explanation for this surprising result is that RNAPII is defective and causes aberrant transcription of *ura4*$^{+}$, attracting chromatin modifiers which allow heterochromatin formation even with defective RNAi. This theory suggests that the RNAi-targeting for heterochromatin formation at the pericentromere is bypassed in the RNAPII mutants. This does not explain why the marker gene is no longer repressed by the pericentromeric heterochromatin however.

Puzzlingly, the RNAPII mutants display defects in silencing of the *ade6*$^{+}$ (Appendix – Figure 1, 2) and *ura4*$^{+}$ (Appendix – Figure 3) marker genes inserted in the outer pericentromeric repeats despite having high levels of H3K9me2 on the repeats (Figure 4.4). This decoupling of silencing and heterochromatin marks is unusual, although this decoupling was also reported for *cid14Δ* cells

(Buhler et al., 2007). Cid14 is a poly(A) polymerase suggested to be required for heterochromatin silencing downstream of heterochromatin assembly. *cid14 deletion* affects heterochromatin integrity at both pericentromeres and the mating type locus. As RNAi mutants only affect pericentromeric silencing, checking silencing at the other heterochromatin loci in the new RNAPII mutants will provide information concerning which stage of silencing the mutants are defective.

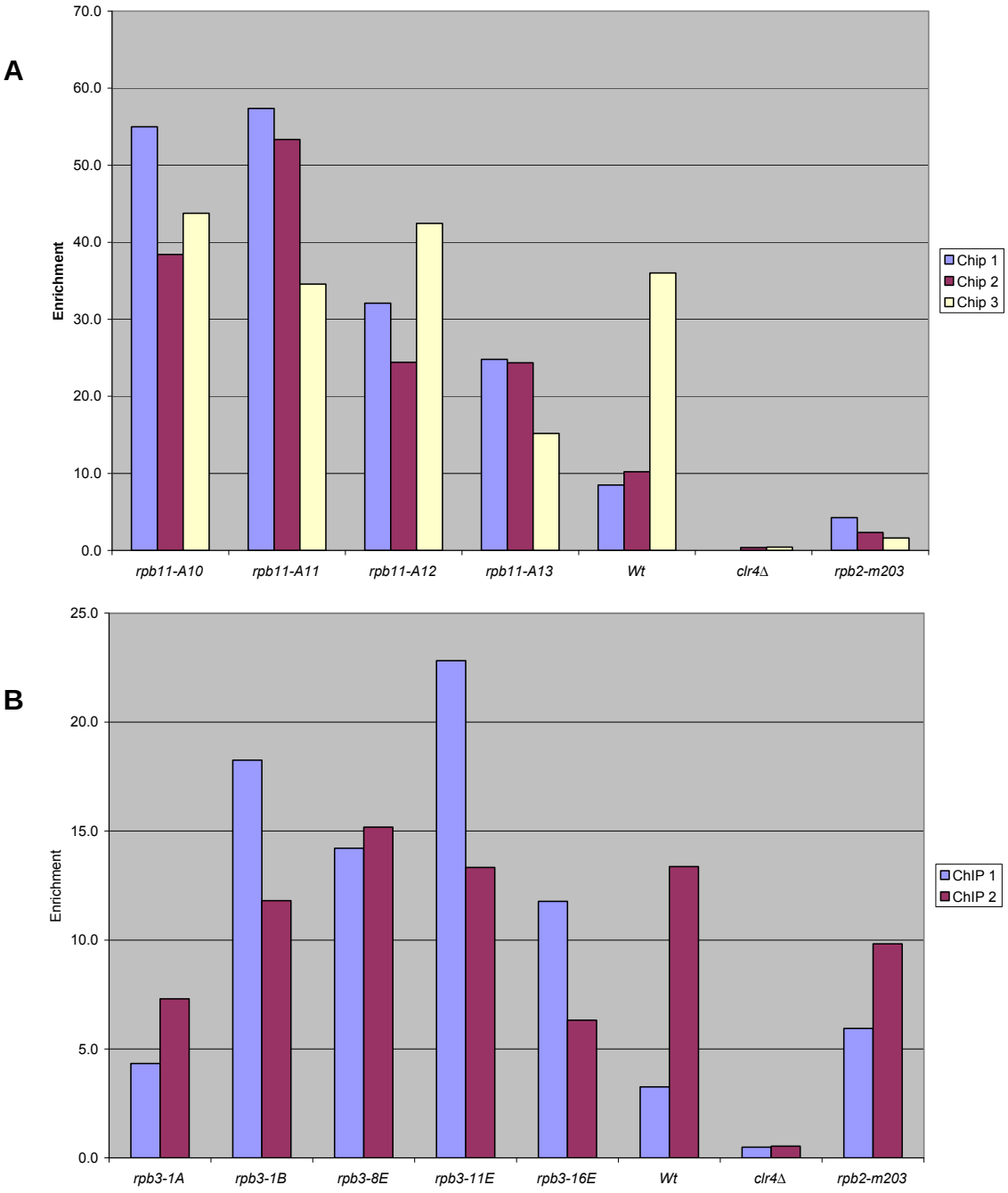
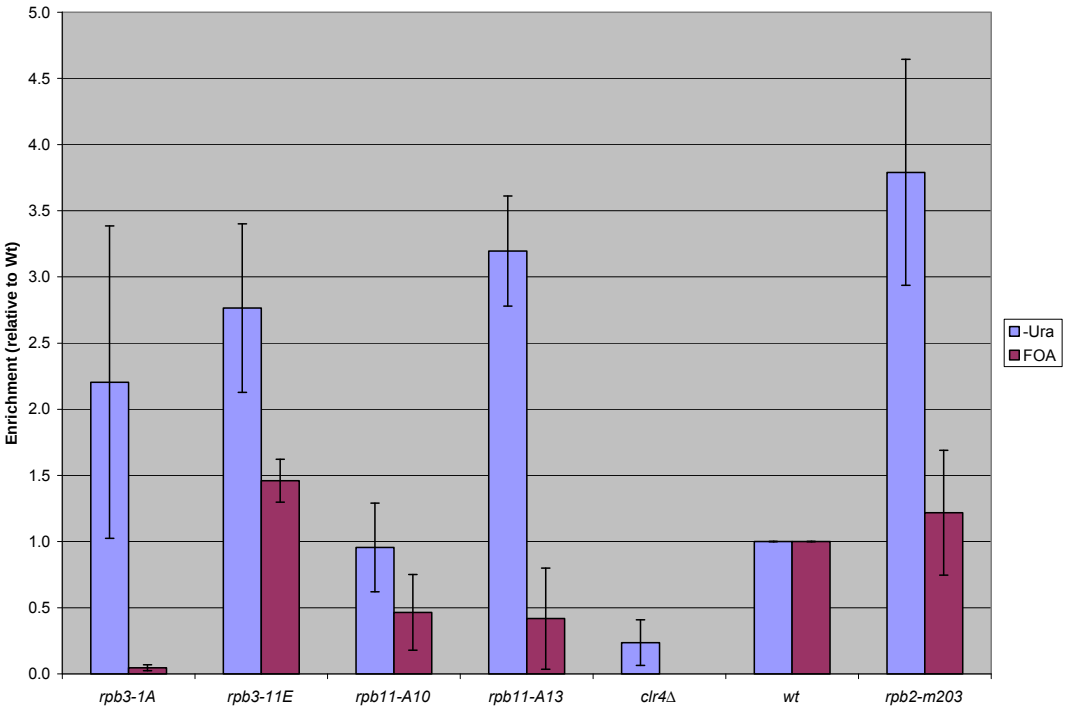


Figure 4.4 – RNAPII mutants retain high levels of the heterochromatin mark H3K9me2. H3K9me2 levels on centromeric dg repeats in the *rpb* mutants. ChIP analysis was used to determine levels of H3K9me2 on cen-dg repeats, measured by qPCR with enrichment values calculated by IP relative to input and normalised using the actin gene. (**A**) *rpb11* mutants. (**B**) *rpb3* mutants.



4.2.3 RNAPII mutants display heterochromatin silencing defects primarily at the pericentromeres

Deletion of any one of the genes of the RNAi pathway results in loss of pericentromeric silencing, but the silencing of a marker gene at telomeres and the mating type locus remain intact. In contrast deletion of the gene responsible for H3K9 methylation (*clr4*) results in complete loss of silencing at all three loci. This is because there is a second pathway that acts redundantly with RNAi to recruit Clr4 to these regions (Atf1/Pcr1 at the mating-type and Taz1 at telomeres). Due

to the difference in silencing at the heterochromatic loci it is important to test the RNAPII mutants for silencing at the other loci as well as the centromeric regions. Serial dilution comparative plating assays were carried out to test silencing of marker genes at the mating type and telomere (Figure 4.6). In addition silencing was tested in the central core of centromere one, which is assembled in unusual chromatin containing the histone H3 variant CENP-A^{Cnp1}. The serial dilution comparative plating assays show that the RNAPII mutants do not have a silencing defect at these other heterochromatic loci or with the central core of cen1. These analyses indicate that the RNAPII mutants are primarily defective in pericentromeric silencing and this suggests that the phenotype is probably due to a defect in the RNAi pathway. In order to ascertain at what part in the RNAi pathway the mutants are defective further experiments are required.

4.2.4 RNAPII mutants – *rpb3-11E* and *rpb11-A10* - induce RNAi-independent silencing at the pericentromere

Recently the Grewal lab presented data on an mRNA export factor, Mlo3, and the RNAPII elongation factor Tfs1^{TFIIS}, in respect to RNAi-dependent heterochromatin function (personal communication R. Allshire). Mlo3 is an RNA binding protein and *mlo3*$^{+}$ overexpression has previously been found to cause chromosome segregation defects (Javerzat et al., 1996). Tfs1 plays a crucial role in releasing stalled RNAPII through retuning the RNAPII active site to cause cleavage of the transcript (Kettenberger et al., 2003), (Kettenberger et al., 2004). In the *mlo3*Δ and *tfs1*Δ mutants, silencing defects normally caused by RNAi null mutants are partially suppressed (unpublished observation Grewal Lab). The best explanation for this phenotype is that the RNAPII complex stalls whilst transcribing pericentromeric repeats, providing a mechanism for recruitment of silencing factors. Since the initial reports on Mlo3 it has been published that *mlo3* mutants display RNAi defects and that Mlo3 is likely involved in suppression of antisense transcripts (Zhang et al., 2011) although the findings described above have not yet been published.

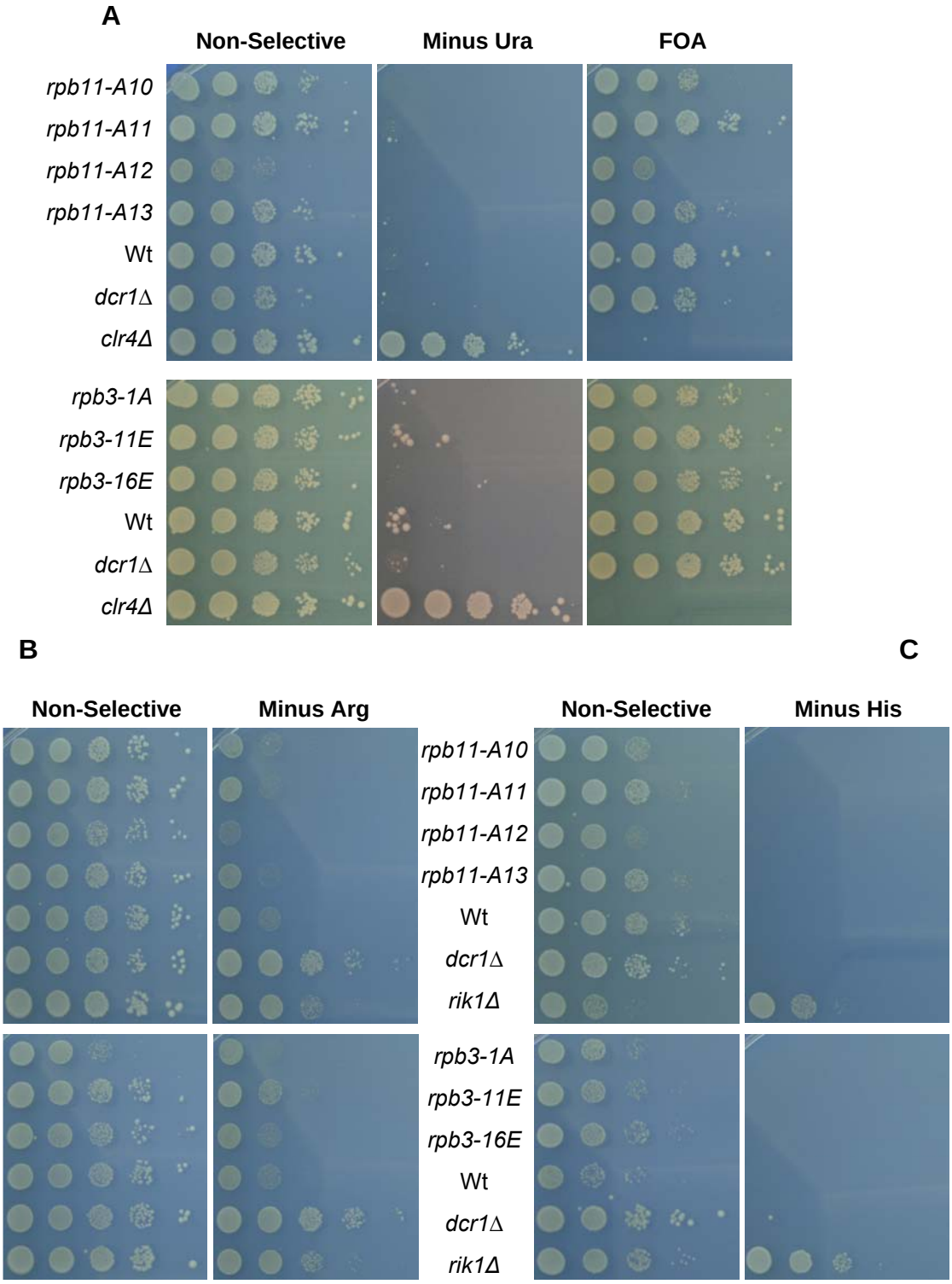


Figure 4.6 – RNAPII mutants display intact heterochromatic silencing at non-pericentromeric loci. Serial dilution comparative plating assays were use to test the silencing of marker genes at different heterochromatic loci. (**A**) The *ura4+* gene at the mating type locus (*mat3*). (**B**) The *arg3+* gene at the central core of centromere 1. (**C**) The *his3+* gene at telomere 1.

As the RNAPII mutants most likely affect transcription of pericentromeric repeats, it is possible that they might have a similar effect. To test this hypothesis silencing of pericentromeric cen1:*ade6*$^{+}$ was assayed in RNAPII-RNAi double mutant backgrounds (Figure 4.7). Surprisingly, both *rpb11-A10* and *rpb3-11E* mutant cells retain the same variegating colony colour even in *dcr1*Δ or *ago1*Δ. Interestingly, *rpb2-m203* does not exhibit this dominant effect over RNAi mutants (Appendix – Figure 5). In order to ensure that the mutants are truly bypassing RNAi to cause silencing, the RNAPII mutants were also assayed for silencing in a *clr4*Δ background (Figure 4.7). As expected, the *rpb11-A10 clr4*Δ and *rpb3-11E clr4*Δ double mutants are both defective for silencing of the marker gene. Therefore, like *mlo3*Δ and *tfs1*Δ, RNAPII mutants can cause partial silencing that is dependent on Clr4 but independent of RNAi. A possible explanation for this result is that the two RNAPII mutants are preventing the RNAPII complex from releasing the transcript resulting in pausing which is causing Clr4-dependent silencing independent of RNAi.

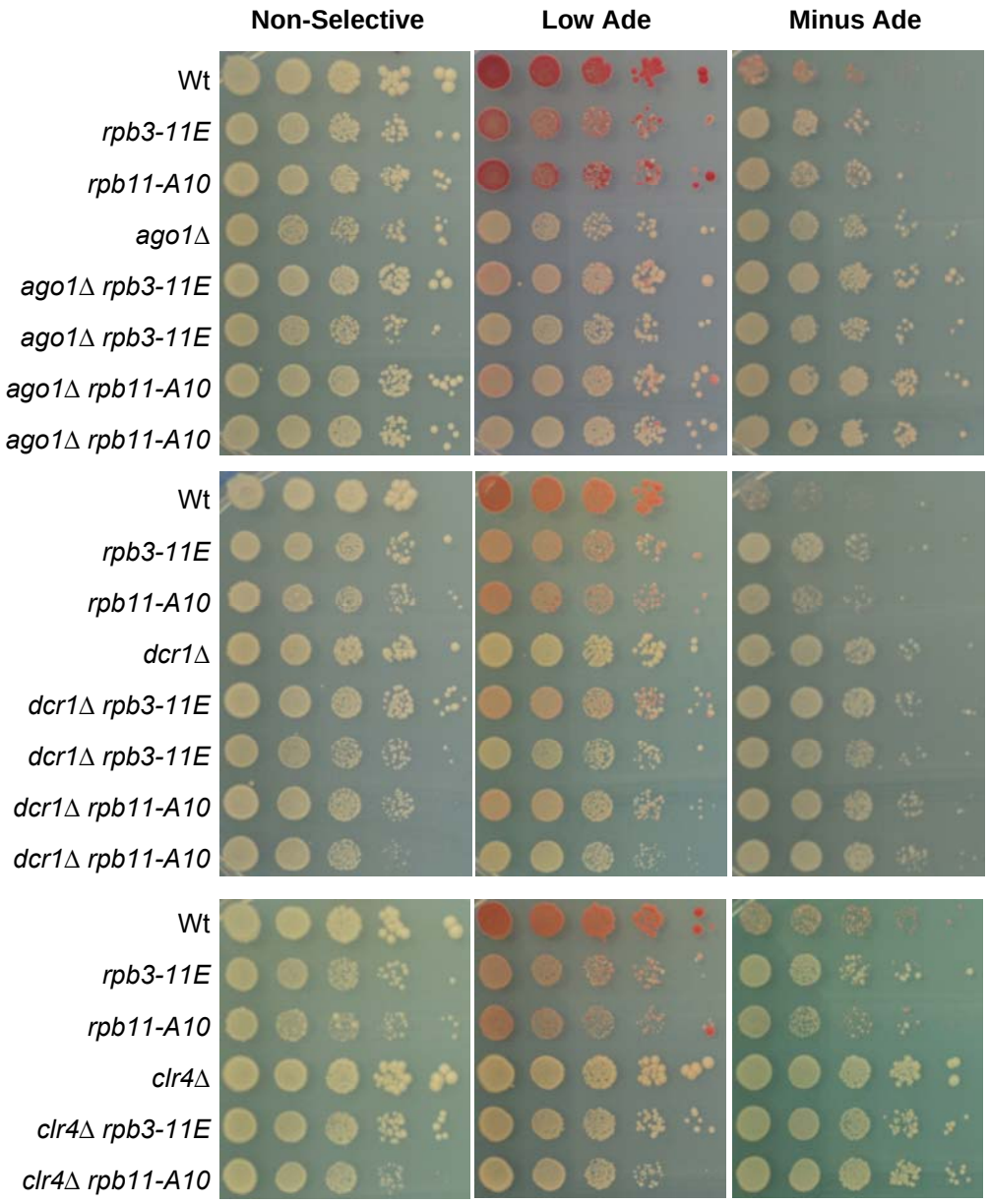


Figure 4.7 – RNAPII mutants partially rescue RNAi silencing defects. Serial dilution comparative plating assays were used to test the silencing of the *ade6+* reporter gene at the pericentromere in RNAPII-RNAi double mutants. Normally the RNAi mutants, *ago1*Δ and *dcr1*Δ, lose silencing of reporter genes at the pericentromere and the ClrC mutant *clr4*Δ loses silencing at all heterochromatin loci. Here *rpb11-A10* and *rpb3-11E* are shown to partially rescue the complete loss of silencing in *dcr1*Δ and *ago1*Δ in a Clr4-dependent manner.

4.3 Discussion

In this chapter a variety of phenotypes were tested for the RNAPII mutants generated in Chapter 3. As RNAPII transcription is absolutely required for siRNA generation, initial analyses focussed on measuring the levels of non-coding centromeric transcript and their resulting siRNAs (Figure 4.1). Many of the mutants displayed reduced siRNA abundance however due to the variability of the siRNA levels observed for each mutant (Table 4.2) additional experiments were designed to compare phenotypic variegation in silencing with siRNA levels. Although phenotypic variegation in centromeric marker gene expression was shown to have a correlation with the siRNA levels (Figure 4.2), it was found that far greater differences in siRNA levels were observed between different isolates of the same RNAPII mutant (Figure 4.3). This inherent variability made it difficult to determine the effects of these mutants on siRNA generation. ChIP analyses showed that the degree of methylation of histone H3 on Lysine 9 (H3K9) of the dg outer repeats of the pericentromere in the RNAPII mutants was relatively unchanged compared to wildtype cells (Figure 4.4). Thus there is an apparent paradox between hallmarks of intact heterochromatin and the observed phenotypic defects in silencing. To test if the mutants were defective in general heterochromatin function, comparative plating assays were used to test the silencing of marker genes at several distinct heterochromatic loci (Figure 4.7). These analyses showed that all of the mutants tested exhibited a specific defect in silencing at pericentromeric outer repeats but, as in RNAi mutants, silencing remained intact at the other loci. This result suggests that RNAPII integrity is specifically important for pericentromeric heterochromatin integrity. In a final experiment to test if these newly isolated RNAPII mutants have similar silencing defects as RNAPII elongation mutants it was found that these RNAPII mutants also partially rescue loss of silencing in cells lacking RNAi components, *dcr1*Δ and *ago1*Δ (Figure 4.7), and that this is dependent on the H3K9 methyltransferase Clr4.

The comparison of centromeric siRNA and transcript levels (Table 4.1) shows that distinct RNAPII mutants are likely to have defects at different stages

in pericentromeric heterochromatin integrity. The three temperature sensitive (ts) *rpb3* mutants; *1A,1B* and *8E* are the slowest growing of the new RNAPII mutants and, excluding *rpb3-8E*, they have the strongest defects in silencing. Their siRNA/transcript levels are quite different, however, with *rpb3-1A* appearing to have approximately wildtype siRNA/transcript levels, whilst *rpb3-1B* and *rpb3-8E* have low levels of both siRNA and transcript accumulation. The low growth rate, ts and siRNA phenotypes observed for *rpb3-1B* and *rpb3-8E* suggest that defects in transcription are causing the disruption of silencing. Three of the faster growing RNAPII mutants; *rpb3-11E*, *rpb11-A10* and *rpb11-A13* appear to have defects in siRNA processing with relatively high levels of transcript accumulation and reduced levels of siRNA. Unfortunately these siRNA processing phenotypes were not consistently detected in the mutants (Table 4.2), so that the results in Table 4.1 are of limited value, as they just represent a snap-shot of a variable cell population with a destabilised RNAi system. Assessment of variegation of marker gene expression on siRNA levels (Figure 4.2) where silent or expressing cells were selected, depending on the growth media, shows a consistent shift in siRNA levels for all mutants examined and even wildtype cells. This correlation between decreased siRNA levels upon selection of marker gene expression is logical. The fact that there is not a larger difference between levels of siRNA for cells grown in selective versus counter-selective media observed for the RNAPII mutants suggests that variegation of heterochromatin assembly/disassembly is not causing the high degree of variability of siRNA levels in different cell extracts.

The variability of the siRNA levels for all of the RNAPII mutants examined is a phenomenon not normally seen in known mutants that lack heterochromatin components (eg *ago1*Δ*, dos1*Δ). This may be because a full gene deletion has a more complete or stable effect on siRNA processing whereas the mutated RNAPII genes have a more subtle and unstable effect. Variability in siRNA levels has been previously observed for the *epe1Δ* mutant (Trewick et al., 2007). Epe1 contains a JmjC domain, a domain known to provide histone demethylase activity in other proteins (Klose et al., 2006) however the JmjC domain of Epe1 does not have demethylase activity *in vitro* (Zofall and Grewal,

2006). It has been suggested instead that Epe1 has protein hydroxylase activity which could disrupt protein interactions or destabilise proteins required for heterochromatin formation (Trewick et al., 2007). Epe1 is important for ensuring all heterochromatin domains are properly defined and in cells lacking *epe1* the extent of heterochromatin becomes more variable, so that it spreads and recedes beyond its normal domains (Trewick et al., 2007). In *epe1Δ* cells the levels of siRNA are proportional to the levels of silencing seen, but as silencing at all heterochromatin domains is affected by *epe1Δ*, the silencing defects are thought to be independent of the RNAi pathway and thus the variability of the siRNA levels does not cause the silencing defect. Similar to *epe1*Δ*,* the pericentromeric silencing in the RNAPII mutants described here is in constant flux, and this is detected as phenotypic variegation. Unlike *epe1Δ,* defective silencing in these RNAPII mutants specifically affects centromere repeats (Figure 4.7); this indicates that the silencing defect is likely due to abrogation of the RNAi pathway. However interpretation of the phenotypes is further complicated by the fact that two of the new RNAPII mutants tested partially rescue silencing of cells lacking RNAi (*dcr1*Δ and *ago1*Δ) (Figure 4.7). Thus it is likely that the RNAPII mutants have more complex affects on silencing within pericentromeric heterochromatin, on the one hand causing significant negative effects through the reduced activity of the RNAi pathway and on the other hand having smaller positive effects by somehow promoting heterochromatin formation independently of RNAi.

There is a seemingly paradoxical difference between the high H3K9me2 levels seen in the RNAPII mutants at the dg repeats (Figure 4.4, Figure 4.5) and the fact that silencing of marker genes is defective in the mutants (Chapter 3 Figure 3.4, 3.5). The only other mutant that displays a similar disconnection of silencing and heterochromatin marks is *cid14*Δ (Buhler et al., 2007) which has intact H3K9me2 at pericentromeric repeats and the mating type locus but is defective in silencing at both loci. Cid14 is a poly(A) polymerase which is one of the proteins that make up the TRAMP complex. It is thought that the TRAMP complex in *S. pombe* mediates degradation of RNA by either the exosome or RNAi and it is believed to act downstream of H3K9 methylation and

heterochromatin protein deposition in silencing of heterochromatin regions (Buhler et al., 2007). The RNAPII mutants have defects in pericentromeric outer repeat silencing only suggesting that the mechanism of this phenomenon is different. The splicing factor mutants described by Bayne et al 2008 also have RNAi-specific silencing defects as they display only a modest reduction in methylation of H3K9 which is specific to the pericentromeric repeats. It is possible that the defects observed here in the RNAPII mutants and previously in splicing mutants have a common source. For example splicing is known to occur co-transcriptionally and is also coordinated with chromatin modifications (Sims et al., 2007), (Luco et al., 2010). It has also been shown that transcriptional pausing affects splicing (de la Mata et al., 2003) and that splicing factors can regulate transcription (Fong and Zhou, 2001), (Kwek et al., 2002).

The ability of *rpb11-A10* and *rpb3-11E* to partially rescue silencing defects of RNAi mutants in a Clr4-dependent manner is quite unexpected (Figure 4.7). This experiment in combination with the *mlo3*Δ and *tfs1*Δ ability to also rescue silencing in RNAi mutants (unpublished observation Grewal Lab) may indicate that formation of heterochromatin and modification of H3K9 involves aberrant elongation by RNAPII. Mlo3 is an RNA binding protein that was first identified in *S. pombe* in a screen for factors causing defects in chromosome segregation upon overexpression (Javerzat et al., 1996), suggesting a possible role in chromosome condensation or centromere/kinetochore assembly. Recently Mlo3 has been identified as the first non-histone protein substrate of Clr4 methylation and it has been found to play a role in RNA transcript identification for degradation (Zhang et al., 2011). Tfs1^{TFIIS} is a well characterised RNAPII general transcription factor required for rescuing stalled RNAPII (Kettenberger et al., 2003), (Kettenberger et al., 2004).

A co-transcriptional RNAi-dependent silencing mechanism has recently been reported in *C. elegans,* this appears to cause RNAPII to pause during elongation and is accompanied by enrichment of H3K9me3 at the RNAi target site (Guang et al., 2010). Although the mechanism of this phenomenon is not yet understood it is possible that RNAi is having a similar effect on transcription in the RNAPII mutants in *S. pombe*. A higher incidence of RNAPII stalling during

elongation may occur in these mutants and this might result in recruitment of Clr4/ClrC. This hypothesis could be initially tested by checking RNAPII occupancy on the reporter gene or the pericentromeric outer repeats in *dcr1*Δ vs *rpb11-A10/dcr1*Δ in order to see if there is RNAPII stalling or defective elongation. As *mlo3*Δ and *tfs1*Δ do not present silencing defects it is unlikely that stalling of RNAPII alone is responsible for the defects in silencing seen in *rpb11-A10* and *rpb3-11E*. This stalling hypothesis may also explain the consistently high levels of H3K9me2 seen on the pericentromeric repeats in the RNAPII mutants.

The fact that *rpb2-m203* does not rescue the loss of silencing in RNAi mutants shows that this is not a phenotype that is generally associated with all RNAPII mutants that affect silencing. This result suggests that there are at least three classes of RNAPII mutants that affect heterochromatin formation in different ways: (1) *rpb2-m203* is likely to be defective in recruiting silencing proteins such as ClrC (2) *rpb7-G150D* is defective in transcription causing reduced levels of RNA substrate for siRNA generation (3) *rpb3-11E* and *rpb11-A10* display RNAi-independent heterochromatin formation at the centromere.

Since all of the genes for the RNAi pathway proteins are transcribed by RNAPII consideration must be given to the possibility that RNAPII mutants affect centromere silencing indirectly through inefficient transcription. This possibility is supported by the observation that all of the RNAPII mutants are slower growing than wildtype suggesting that genes required for optimum cell growth are not being normally expressed. Ultimately this could be addressed by checking genome-wide transcript levels by micro-array analyses for each of the mutants, as previously performed for both the *rpb7-G150D* (Djupedal, 2005) and *rpb2-m203* (Kato, 2005) mutants.

All of the RNAPII mutants tested displayed variable levels of pericentromeric siRNA, including the previously published mutant *rpb2-m203*. The experiments carried out so far have not been sufficient to determine if the variable siRNA levels are the source of the silencing defects or simply a consequence of them. All of the mutants, including *rpb2-m203*, also have variable but essentially normal levels of H3K9me2 on pericentromeric repeats. There does not appear to be a clear correlation between the severity of the silencing

defects, the levels of siRNA or the levels of H3K9me2 for any of the mutants examined. This lack of consistency between the important phenotypes associated with pericentromeric heterochromatin integrity may well explain why no further analyses of *rpb2-m203* have been published to address the roles that RNAPII plays in this system. The inconsistencies also suggest that RNAPII may have a key role at numerous stages starting with transcription of the pericentromeric repeats but also for other downstream events on the way to silencing. The ability of *rpb3-11E* and *rpb11-A10* to partially rescue silencing independently of RNAi demonstrates that the role of RNAPII in forming heterochromatin is much more intricate than might at first have been realised.

Chapter 5 - Identification of the affected gene in *csp6* mutants which are defective in pericentromeric silencing

5.1 Introduction

A collection of mutants were previously isolated in a screen designed to identify factors involved in the formation of heterochromatin at the pericentromere (Ekwall, 1999). The screen involved random mutagenesis, using Ethyl Methanesulfanate (EMS) which introduces point mutations, of a tester strain which had the reporter gene *ade6*$^+$ inserted into the heterochromatin repeats of the pericentromere. The *ade6*$^+$ gene encodes a member of the AIR (amino-imidazole ribonucleotide) Carboxylase family (Fisher, 1969) an enzyme required in the purine biosynthesis pathway. When expression of the *ade6*$^+$ gene is compromised the upstream substrate for CAIR accumulates (when the cells are grown on plates with low adenine) which when oxidised results in red colony colour. Thus when inserted within heterochromatin the *ade6*$^+$ gene is silenced and red colonies are formed. A defect in heterochromatin allows expression of this *ade6*$^+$ gene so that the substrate is converted to CAIR and white colonies are formed. The original screen resulted in the isolation of 13 mutants (representing 12 individual loci) which were designated *csp 1* to *13* for *c*entromere: *s*uppressor of *p*osition effect (Table 5.1). The *csp1* to *6* mutants are temperature sensitive indicating that they encode critical proteins whereas *csp7* to *13* are not temperature sensitive.

In the ts class the affected gene has been identified for *csp3* (*rpb7* – RNAPII subunit), *csp4* (*cwf10* – splicing factor) and *csp5* (*prp39* – splicing factor). These ts mutants cause mutations in proteins that have been well characterised and are known to have roles in essential pathways such as transcription and splicing, both of which have been confirmed to impact on RNAi-directed heterochromatin formation (Djupedal, 2005), (Bayne et al., 2008). In the non-ts class the affected genes encode RdRP (*csp7*), Cid12 (*csp8*), Ago1 (*csp9*) and Arb1 (*csp12*). The only known role for these proteins is in the RNAi pathway. In the original screen two alleles of *csp6* were isolated. Since both mutants had strong

effects on silencing (Ekwall, 1999), siRNA levels and H3K9 methylation (Table 5.2) (Portoso, 2005) its identity is of particular interest as it must play a critical role in pericentromeric heterochromatin formation.

ts Mutants		
Mutant	**Gene Identified**	**Role in *S. pombe***
csp1	***---------?---------***	
csp2	***---------?---------***	
csp3	rpb7	RNAPII Subunit
csp4	cwf10	Splicing Factor
csp5	prp39	Splicing Factor
csp6	***---------?---------***	

non-ts Mutants		
Mutant	**Gene Identified**	**Role in *S. pombe***
csp7	rdp1	RNAi Component
csp8	cid12	RNAi Component
csp9	ago1	RNAi Component
csp10	cid12	RNAi Component
csp11	***---------?---------***	
csp12	arb1	RNAi Component
csp13	***---------?---------***	

Table 5.1 – A list of the *csp* mutants and the known identities of the genes affected

Previous attempts to identify *csp6* have resulted in the gene being mapped to chromosome III (Portoso, 2005) however further mapping has not been successful (Bijos, 2007). Whole-genome plasmid libraries have been used on three separate occasions to identify high-copy suppressors of *csp6* by rescuing the ts phenotype (Ekwall unpublished), (Portoso, 2005). These attempts consistently identified three different Hsp70 genes; *ssa1*$^{+}$, *ssa2*$^{+}$ and *ssp1*$^{+}$. However genetic analyses indicated that these three genes were high copy extragenic suppressors so that the identity of the *csp6* gene remained elusive.

	Silencing Defect*	Growth on TBZ	H3K9me2 at Cen	Swi6 Localization	Cen Transcript Levels	Cen siRNA Levels
wt	-	+++	+++	+++	-	++
csp1	-	+++	+++	+++	+	-
csp2	-	+	+++	+++	+	++
csp6	++	-	-	+	-	-
*clr4*Δ	++++	-	-	-	++	-
*dcr1*Δ	+++	-	+	+	++	-

Table 5.2 – Summary of the phenotypes of the unidentified ts *csp* mutants.
Silencing defects measured using arbitrary scoring of colour of colonies seen by comparative plating assays using strains containing the ade6 marker gene at the pericentromere. TBZ is a microtubule destabilising drug for which pericentromeric heterochromatin defective mutants have sensitivity (arbitrary score based on comparative plating growth assay). H3K9me2 and Swi6 localisation are the hallmarks of heterochromatin (arbitrary scores based on ChIP and immunostaining assays respectively). Pericentromeric repeat DNA is transcribed to form pre-siRNA which is processed into siRNA. Centromeric siRNA is used to target heterochromatin formation to the pericentromere. Arbitrary scoring of centromeric transcript and siRNA levels using northern analyses.
*Silencing defects specifically at the pericentromere region
Adapted from (Portoso, 2005), Chapter 3 Table 1

The heat shock protein (Hsp)70 family of proteins is very highly conserved in almost all forms of life (Richter et al., 2010). Hsp70 proteins are named for their role in heat-shock response found through their up-regulation when the cell is stressed. These proteins act as chaperones for unfolded proteins in order to prevent them from forming aggregates and in order to aid their folding to the native state (Hartl and Hayer-Hartl, 2002). They are also involved in a number of cell processes such as; translocation of proteins across membranes (Tomkiewicz et al., 2007), regulation of signalling pathways (Pratt and Toft, 2003), clathrin uncoating from vesicles (Young et al., 2003) and regulation of protein degradation (Luo et al., 2009). Recently a new role for Hsp70, in loading siRNA into the Argonaute protein, was described (Iki et al., 2010), (Iwasaki et al., 2010), (Miyoshi et al., 2010).

In this chapter I present my characterisation of the *csp6* mutants. Through the application of next generation sequencing I have identified a mutation in

ssa2, which encodes a Hsp70, in both alleles of *csp6* and confirmed that these mutations cause defective heterochromatin formation.

5.2 Results

5.2.1 Characterising the *csp6* alleles

In order to familiarise myself with the phenotypes of both the *csp6-75* and *csp6-95* alleles, I repeated a number of the assays that had been previously performed (Ekwall, 1999), (Portoso, 2005). This also allowed me to carefully compare both *csp6* alleles, since previous tests were carried out along with numerous other *csp* mutants and often included only a single *csp6* allele.

I initially performed serial dilution growth assays to determine the strength of the silencing defects in *csp6-75* and *csp6-95*, and their sensitivity to temperature and to Thiabendazole (TBZ) (Figure 5.1). Silencing defects were assessed in the same way as in the original *csp* screen, using the cen1:*ade6*$^{+}$ reporter gene positioned in the dg outer repeats flanking, with *ade6*$^{+}$ expression levels determined by growing the strains on either low adenine or minus adenine plates. TBZ is a microtubule destabilising drug, thus mutants with a defect in centromere function frequently exhibit increased sensitivity to elevated levels of TBZ. Pericentromeric heterochromatin is known to recruit a high density of Cohesin which provides tight sister chromatid cohesion. Failure to recruit Cohesin results in premature centromere separation and an increased rate of chromosome loss. Thus mutants affecting RNAi or key heterochromatin components are sensitive to TBZ. Plates used for testing temperature sensitivity were supplemented with Phloxin B, which stains dead cells red.

Silencing of the pericentromeric *ade6*$^{+}$ gene was lost as both mutants allowed increased growth on minus adenine plates but only *csp6-95* formed white colonies when grown on low adenine. Under the conditions tested *csp6-95* was clearly TBZ sensitive whereas *csp6-75* showed no sensitivity. These analyses suggest that *csp6-95* has the more extreme phenotypes; the stronger silencing

defects, the slower grower growth, the higher temperature sensitivity and greater TBZ sensitivity. Subsequent analysis of levels of centromeric transcript accumulation and siRNAs in the two alleles reveals a similar pattern. RT-PCR analysis shows that at both permissive and restrictive temperature *csp6-75* cells accumulate more centromeric transcript than wildtype cells, but not as much as *csp6-95* cells. Correspondingly, analysis of siRNA levels indicates that siRNA accumulation is modestly reduced in *csp6-75* cells, but much more strongly affected in *csp6-95* cells. Thus *csp6-95* appears to be a more strongly defective allele than *csp6-75*. (Figure 5.2).

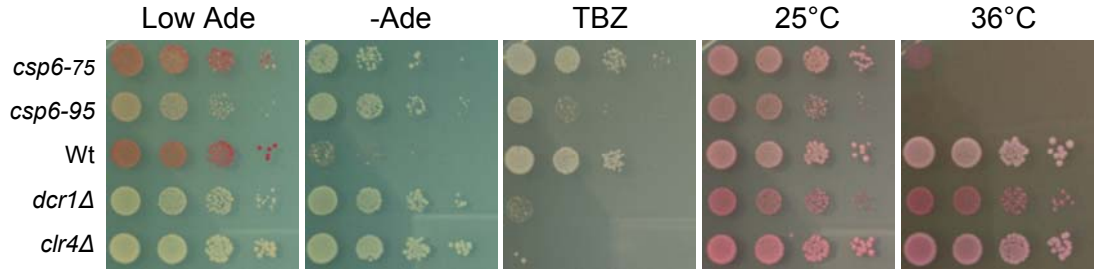


Figure 5.1 – Silencing defects and temperature sensitivity of the *csp6* alleles. Serial dilution comparative plating analyses were used to test the phenotypes. The *ade6*$^{+}$ gene, which is situated in the pericentromeric heterochromatin region, allows silencing to be tested using growth on low adenine (wt is red, defective silencing is white) and minus adenine (no growth for wildtype, good growth for defective silencing) plates. Cells with defective pericentromeric heterochromatin have defects in cohesion between sister chromatids resulting in sensitivity to the microtubule destabilising drug TBZ. To test temperature sensitivity (ts) cells are grown at permissive (25°C) and restrictive (36°C) temperatures on plates supplemented with Phloxin B which stains dead cells red.

Previous analyses of H3K9 methylation levels associated with pericentromeric repeats in *csp6-95* cells appeared to show complete loss of this post-translational modification, similar to levels in *clr4*Δ cells (Portoso, 2005). However in the Chromatin IP (ChIP) analyses presented here (Figure 5.3), *csp6-95* cells consistently show levels of pericentromeric H3K9me2 that are at least as high as those in wildtype cells.

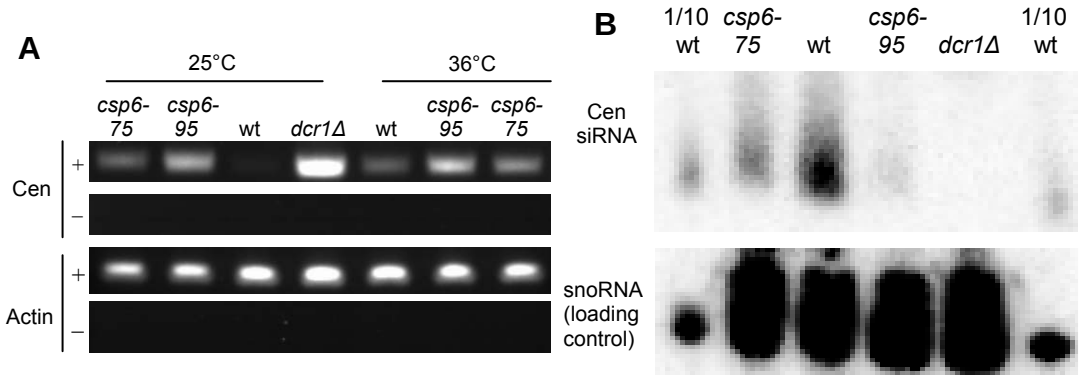


Figure 5.2 – Pericentromeric transcript processing into siRNA in the *csp6* alleles. (A) Centromeric transcript levels determined by RT-PCR at the permissive and restrictive temperatures. For transcript levels at the restrictive temperature the culture was grown at 25°C overnight and then shifted to 36°C for 6 hours. (B) Centromeric siRNA levels, at 25°C, displayed using northern analysis.

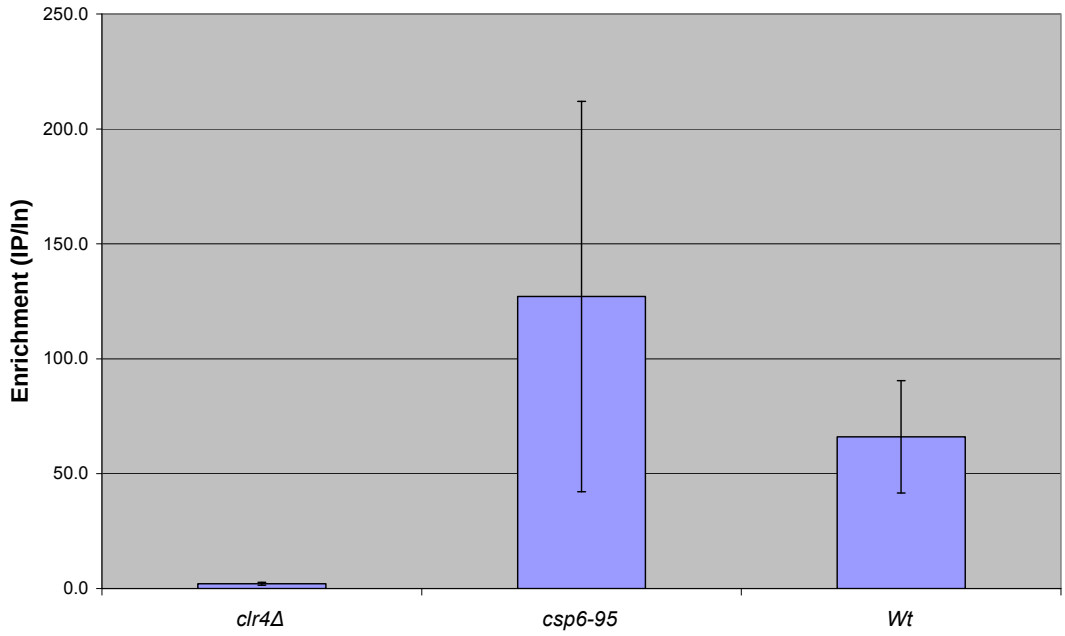


Figure 5.3 – *csp6-95* retains high levels of H3K9me2 on the pericentromeric dg repeats. ChIP was carried out on the samples in triplicate using α-H3K9me2 antibody. Real Time(RT)-PCR was used to quantify dg and actin DNA and the actin value was used to normalise the dg value before calculating enrichment using IP/Input.

The identification of *csp4* and *csp5* as alleles of genes encoding splicing factors (Bayne et al., 2008) and the association of splicing factors with RdRC (Motamedi et al., 2004) has led to the discovery that specific splicing factors play a role in the formation of heterochromatin. Like *csp6-95*, some splicing factor mutants also retain relatively high levels of centromeric H3K9me, despite generating only low levels of siRNAs. To determine if the *csp6* alleles affect splicing, a simple splicing assay for splicing efficiency was employed (Figure 5.4). This RT-PCR-based assay uses primer pairs positioned in separate exons of a gene to assess the proportions of RNA which either retain the intron (unspliced pre-mRNA), or lack the intron (spliced mature mRNA). Splicing of the *ade9*$^+$, *tbp1*$^+$ (TFIID) and *cdc2*$^+$ genes was tested. These genes were chosen because *cdc2*$^+$ and *tbp1*$^+$ were used previously to characterise splicing defects (Habara et al., 2001) and the *ade9*$^+$ gene was used in the initial characterisation of *csp4* and *csp5*. No increased accumulation of unspliced transcript was observed in either *csp6-75* or *csp6-95* grown at 25°C and 36°C however, mutation in the *prp1*$^+$ gene (a known splicing factor) clearly accumulate pre-mRNA relative to wildtype cells (Figure 5.4). These results indicate that the *csp6* alleles both have fully functional splicing at both permissive and restrictive temperatures; this indicates that *csp6*$^+$ probably does not encode a splicing factor.

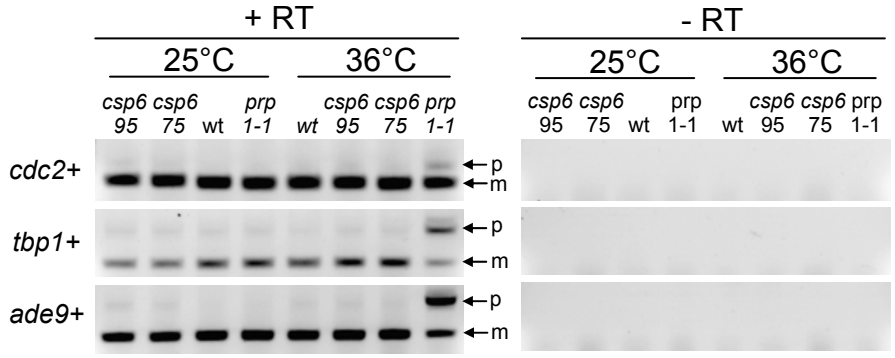


Figure 5.4 – The *csp6* mutants do not display splicing defects. A splicing assay using reverse transcribed RNA was used to assess if introns are properly spliced. Primers were chosen from separate exons in order to see if the intervening intron was efficiently spliced out. *cdc2+* and *tbp1+* have been used to characterise splicing defects previously (Habara et al., 2001). *ade9+* is an additional intron containing gene which was used in initial characterisation of *csp4* and *csp5*. 'p' denotes the pre-mRNA (unspliced) whilst 'm' denotes the mature mRNA (spliced). The *prp1-1* ts splicing mutant has defective splicing at 36°C and so acts as a positive control.

5.2.2 Identification of the *csp6*$^{+}$ gene by genome resequencing of *csp6-75* and *csp6-95*

Previous attempts to identify the affected gene in the *csp6-75* and *csp6-95* mutants using genetic linkage analysis indicated that the mutations were linked to chromosome III, but further mapping and complementation with genomic libraries failed to determine the identity of the gene affected.

The recent introduction of next generation sequencing platforms means that genome resequencing has become a powerful and affordable tool in identifying causative mutations. This technique has been used to identify causative mutations in *D. melanogaster* (Blumenstiel et al., 2009), *C. elegans* (Sarin et al., 2008), and *S. pombe* (Irvine et al., 2009).

In order to identify the affected gene the Illumina GAII platform was used to sequence DNA fragments of 35 bp in length. Initially *csp1, csp2, csp6-95, csp11, csp13* and a wildtype strain were chosen for sequencing in parallel. The advantage of sequencing these strains at the same time is that, because the strains are known to have genetically distinct causative mutations, any single nucleotide polymorphisms (SNPs) introduced during strain construction and shared by

more than one of the strains can be discarded. When this project began this genome resequencing approach was unpublished and so we decided to use one Illumina flow cell lane per sample in order to get 8-10 fold coverage of the genome (ie a mean of 8-10 reads per base). A subsequent study in *S. pombe* (Irvine et al., 2009) has suggested that 20-25 fold coverage of the genome should be adequate for identification of candidate mutations for further testing.

As the *csp* mutants were generated by random mutagenesis using EMS it is likely that multiple distinct mutations occurred in each mutant strain, although the exact mutation rate is unknown. Three backcrosses were originally carried out to dilute out potential mutations unlinked to the mutation that causes defective silencing. In order to further reduce the number of background mutations detected by genome resequencing two additional backcrosses were carried out. Ten independent colonies were isolated from the final backcross and were grown in parallel and mixed in equal cell numbers before being harvested for DNA extraction. The use of multiple independent isolates means that the causative mutation and closely linked background mutations are the only SNPs likely to be present in 100% of the sequencing reads for a given base.

The results of the sequencing are summarised in Table 5.3. Initially the sequencing data was analysed using very stringent filters where SNPs have to meet the following parameters: (1) base appears mutated in >90% of the reads (2) base must have >5 good quality reads. These filters ruled out all of the SNPs for *csp13,* and only one mutation was left on chromosome III for *csp6-95,* which was not associated with an Open Reading Frame (ORF). The analysis was then repeated using a less stringent set of filters: (1) base appears mutated in >70% of reads (2) base must have >3 good quality reads. These filters ruled out all of the SNPs on chromosome III for *csp6* because the reduced filtering allowed more of the 70-90% mutated base calls from all of the strains sequenced to be counted, some of which cancelled out previously unique hits. Two SNPs for *csp6,* five SNPs for *csp1* and two SNPs for *csp2* were tested by transformation with wildtype DNA fragments but none of the tested fragments rescued the ts and silencing

defects. Further analyses of *csp11* and *csp13* by resequencing (S. White, Allshire Lab) also proved unsuccessful.

	Genome Coverage		High Stringency	Reduced Stringency
Strain	% of Genome <5 reads	Mean reads per base	Unique SNPs when residue mutated in >90% of reads	Unique SNPs when residue mutated in >70% of reads
csp1	8.21	12.26	6	32
csp2	0.14	29.19	3	24
csp6-95	0.43	28.20	4	17
csp11	9.16	11.49	1	25
csp13	7.54	12.09	0	25
Wt	19.68	9.41	ND	ND

Table 5.3 – The results of the first Illumina genome resequencing attempt to identify the remaining *csp* mutants.

Although the average number of reads per base were relatively high, and comparable to the 20-25 fold coverage recommended by Irvine et al 2009, the actual percentage of the genome that was under-represented was significant (Table 5.3). We therefore decided to repeat the sequencing, ensuring that the percentage of the genome with low coverage would be below 0.1%, and with the SNP data filters set to (1) base mutated in >90% of the reads (2) base must have at least 3 good quality reads. Improved coverage was made possible by new advances in Illumina GA technology which meant that we could obtain longer sequence reads (~50 bp). We also used paired-end reads which improve the mapping quality of the reads to the reference sequence as well as costing less per base than the single-end reads. To increase our chances of identifying the affected gene in *csp6* mutants we chose to sequence the genomes of both *csp6-75* and *csp6-95*. Sequencing both *csp6* alleles would allow us to search for mutations in the same gene in these independent strains, allowing background mutations to be filtered out and thus making the affected gene easier to identify.

The second set of sequencing data proved to be much more comprehensive, producing >80 fold coverage of both genomes and reducing the percent of low coverage to ~0.1% (Table 5.4). Just two unique SNPs were

identified for *csp6-95* whilst sequencing of *csp6-75* revealed 12 SNPs which differ from both the reference and the previously sequenced mutants. Comparison of the unique SNPs in both alleles showed that *ssa2*$^+$, which encodes a Hsp70 protein, was the only gene harboring a mutation in both strains (Table 5.5). This gene is found on chromosome III thus correlating with the linkage analysis that mapped the *csp6* mutation to chromosome III (Portoso, 2005). This gene is also one of the three Hsp70 genes which had previously been shown to rescue the temperature sensitive phenotype of *csp6* mutations when transformed on high copy plasmids (Portoso, 2005), (Ekwall, 1999). The mutations in the *ssa2*$^+$ gene both affect the same codon resulting in the substitution of a highly conserved alanine residue which resides within the Nucleotide Binding Domain (NBD) of the Hsp70 protein.

Strain	***csp6-95***	***csp6-75***
Mean Reads per base	82.96	97.70
% zero coverage	0.019	0.029
% low coverage (<3)	0.025	0.150
Total Paired End Reads	23,611,174	24,833,034
% reads mapped	99.46	98.28
% mapped in pairs	98.81	78.34
Unique SNPs	2	12

Table 5.4 - The results of the second Illumina genome resequencing attempt to identify *csp6*

Strain	**Chromosome**	**Position**	**Ref**	**Mutant**	**Depth**	**Gene**	**Codon**	**AA**
csp6-95	III	2057397	C	T	68	*ssa2*$^+$	gct-gtt	A52V
csp6-75	III	2057396	G	A	53	*ssa2*$^+$	gct-act	A52T

Table 5.5 - Identification of *ssa2*$^+$ as the mutated gene in *csp6* by whole-genome sequencing shows that the *csp6* alleles both bear mutations in the same codon.

In order to confirm that the mutations in *ssa2*$^+$ are responsible for the ts and silencing phenotypes, two simple tests were performed. Firstly, the *ssa2-A52V* mutant was recreated by co-transforming a wildtype strain with a 300 bp fragment containing the *ssa2-A52V* (C155T) mutation identified in *csp6-95* and a *leu2*$^+$ bearing plasmid. Transformants were selected on minus LEU plates and replica plated to low adenine plates grown at the permissive temperature (25°C). Eight transformants displaying defective silencing of the *ade6*$^+$ marker gene (white colonies) were selected along with two wildtype (red) colonies. Sequencing confirmed that the *ssa2-A52V* (C155T) mutation was only present in the colonies displaying a silencing defect. The phenotypes of these reconstructed mutations were indistinguishable from those of the original *csp6-95* mutant; both strains exhibited a similar defect in silencing as indicated by the change in colony colour on low adenine plates and the growth of colonies on minus adenine plates (Figure 5.5). Moreover the regenerated *ssa2-A52V* mutant exhibited sensitivity to TBZ and temperature sensitive growth at 36°C. This suggests that the mutation is sufficient to account for the observed phenotypes.

Secondly, if the *csp6-95* phenotypes are caused by mutation of the *ssa2*$^+$ gene then substituting the mutated residue with the wildtype sequence should allow repair of the defect in a proportion of transformants. Therefore the *csp6-95* strain was transformed with a 300 bp fragment containing the *ssa2*$^+$ sequence at the mutation site. Transformants were grown at the restrictive temperature (36°C) and the resulting colonies were sequenced for the *ssa2-A52V* (C155T) mutation. Sequencing confirmed that the affected nucleotide had been substituted for the wildtype sequence in all eight transformants tested. This *ssa2*$^+$ fragment was sufficient to fully rescue all the *csp6-95* mutant phenotypes tested (Figure 5.5), confirming that the *ssa2-A52V* allele is the cause of the ts and silencing defects seen in *csp6-95*.

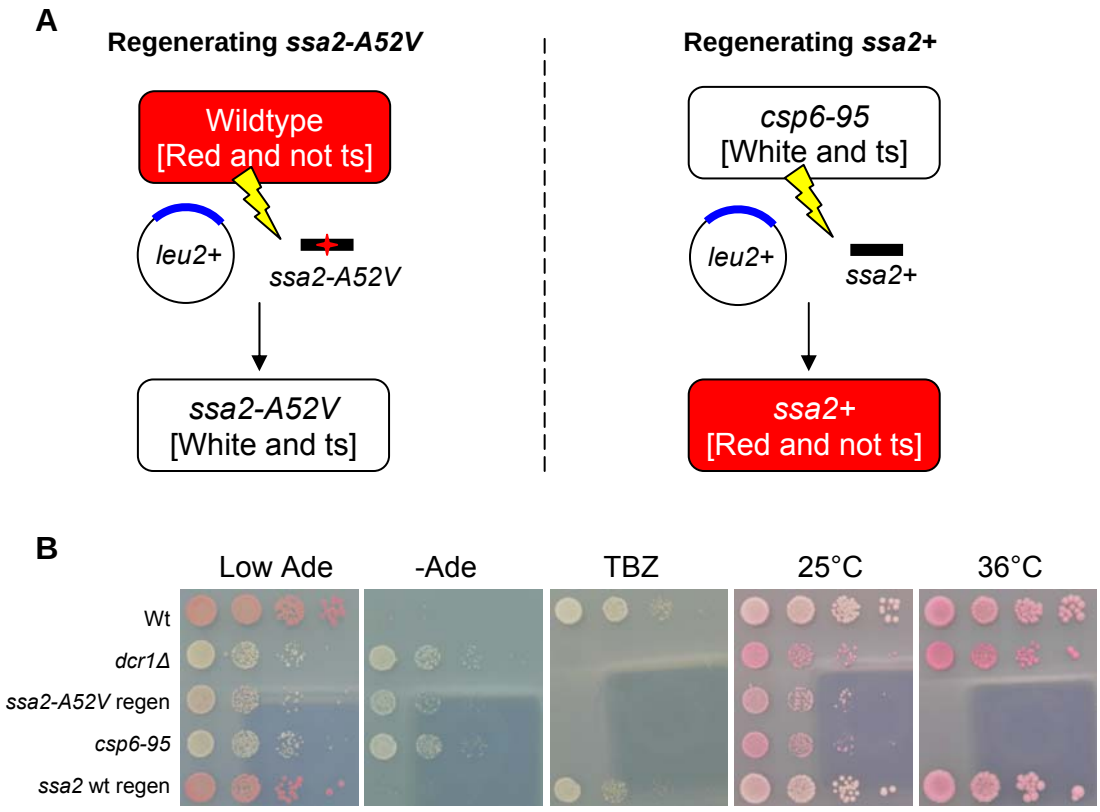


Figure 5.5 – The *ssa2-A52V* mutation causes the temperature sensitivity and heterochromatin phenotypes in *csp6-95*.
(**A**) Illustration of the strategy used to regenerate the *ssa2-A52V* mutation in a wildtype strain and to regenerate *ssa2+* in the *csp6-95* strain. A 300 bp fragment of DNA surrounding the mutation site was co-transformed with a plasmid bearing the *leu2+* gene. Transformants were selected on –LEU plates and replica plated onto plates in order to test the silencing and ts phenotypes. Sequencing was used to confirm nucleotide substitution at the mutation site.
(**B**) Comparative plating assay confirming that the *csp6-95* mutant's ts and silencing defects are caused by the *ssa2-A52V* mutation. This assay tests the silencing of the *ade6+* reporter gene at the pericentromere (low Ade and minus Ade), temperature sensitivity (25°C vs 36°C) and sensitivity to the microtubule destabilising drug TBZ (TBZ) in the new *ssa2* allele strains.

5.2.3 The *csp6* mutations substitute a highly conserved residue in the Hsp70 gene *ssa2*$^{+}$

In *S. pombe* there are two cytoplasmic Hsp70 proteins named Ssa1 and Ssa2. It is known that knockouts of these proteins are viable (Kim et al., 2010) although a double-knockout has not been published in *S. pombe*. In *S. cerevisiae* there are four

cytoplasmic Hsp70 proteins called SSA1, SSA2, SSA3 and SSA4. Knockout of any three out of four of these proteins is still viable, although the SSA3 protein alone requires expression under a stronger promoter, however knockout of all four is lethal (Werner-Washburne et al., 1987). As the *csp6* mutations cause temperature sensitivity and silencing defects it is likely that Ssa2 function is not entirely redundant. In higher eukaryotes there are many more Hsp70 proteins and they have become more specialised. For example in humans and mice HSPA1L is up-regulated during stress whereas HSPA8 is a constitutively expressed Hsp70 (Tavaria et al., 1996). Alignment of Ssa2 with other eukaryotic Hsp70 proteins highlights the extreme conservation of the family (Figure 5.6).

As the *csp6-75* and *csp6-95* mutants have been confirmed to be alleles of the *ssa2*$^{+}$ gene I will now refer to them as *ssa2-75* and *ssa2-95* respectively. To better understand the affect if the *ssa2* mutations on Hsp70 function, the affected residue was mapped onto a 3D Hsp70 structure. The crystal structure of the E. coli Hsp70 homologue DnaK is the most complete example of a Hsp70 3D structure (unpublished Bertelson et al). Both *ssa2* mutants result in a substitution of a very highly conserved alanine residue (Figure 5.6) in the N-terminal Nucleotide Binding Domain (NBD) (Figure 5.7). Figure 5.7 depicts the ADP-bound conformation of Hsp70 which tightly binds substrate peptides in the SBD. Upon nucleotide exchange from the ADP to ATP binding there is a conformational shift translated through the NBD into the SBD causing the lid subdomain to open away from the SBD allowing release of the substrate protein (Jiang et al., 2006).

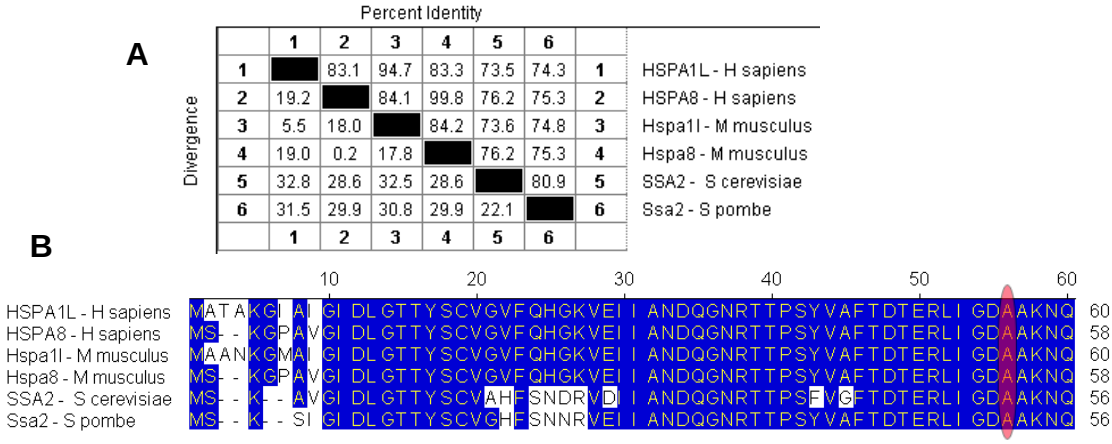


Figure 5.6 – Ssa2 is part of the highly conserved Hsp70 family. This comparison of the heat-induced (HPA1L) and constitutively expressed (HSPA8) Hsp70 proteins from human and mouse with cytoplasmic Hsp70 from yeast used the Jotun Hein Method for alignment. (**A**) Comparison of percentage sequence identity of the Hsp70 proteins. (**B**) The alignment of this N-terminal region of the Hsp70s shows the highly conserved alanine residue, highlighted in red, which is substituted in both of the *csp6* mutant alleles.

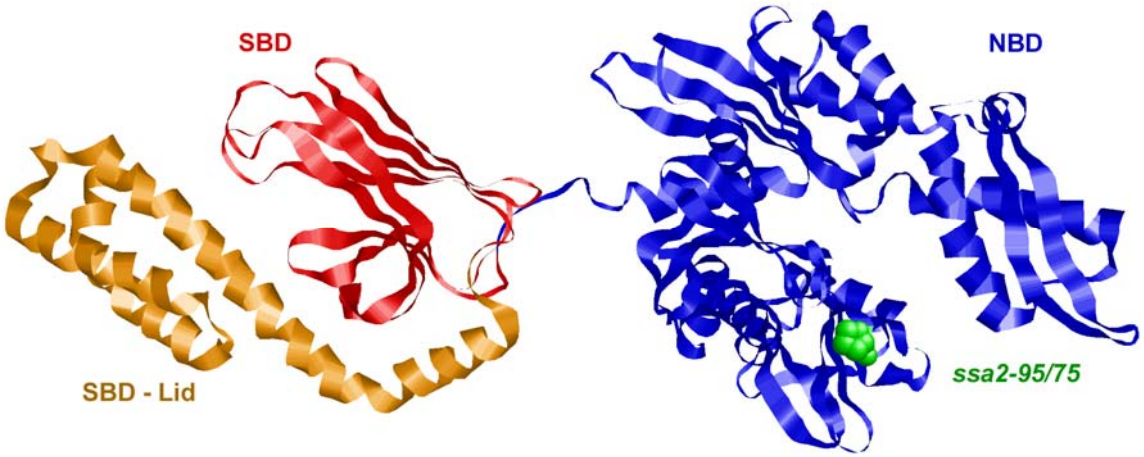


Figure 5.7 – The crystal structure of a Heat-shock protein (Hsp)70 highlighting the position of the residue mutated in *ssa2-75* and *ssa2-95*. The Hsp70 family of proteins has a highly conserved 3° structure made up of an N-terminal Nucleotide Binding Domain (NBD), a Substrate/Peptide Binding Domain (SBD) and a C-terminal 'lid' subdomain of the SBD.
* Crystal structure of the E. coli Hsp70 homolog DnaK – 2KHO – Bertelson et al Unpublished.

The more severely affected allele, *ssa2-95*, has an alanine to valine substitution which is unlikely to disrupt the 2° structure due to the similarity of

the side-chains. It is therefore unlikely that the *csp6* mutants are disrupting the overall stability of the protein. The *ssa2-95* mutation is also well clear of the ATP/ADP interaction site, which is deep inside the cleft formed by the two lobes of NBD, so nucleotide binding is still likely to be intact.

In addition to ATPase activity, the NBD is also important for regulation of the function of the Hsp70 through protein-protein interactions with cofactors. J proteins and Nucleotide Exchange Factors (NEFs) interact with the NBD. J proteins are highly diverse proteins which share a common J domain. The diversity of J proteins is thought to direct Hsp70 protein function, providing the substrate-binding specificity needed for the many roles of Hsp70 (Kampinga and Craig, 2010). The NEFs are evolutionary convergent proteins thought to regulate the speed at which Hsp70 binds and releases substrate proteins using individually unique mechanistic approaches (Cyr, 2008). It seems likely therefore that the substitution of the highly conserved alanine residue in the *csp6* mutants may result in faulty regulation of Ssa2 through one or more of its cofactor proteins.

5.3 Discussion

In this chapter the phenotypes relevant to RNAi-dependent heterochromatin formation of the two *csp6* mutant alleles, *csp6-75* and *csp6-95*, were characterised. The affected gene in *csp6* mutants was also identified to be the heat-shock protein (Hsp)70 coding gene *ssa2*$^{+}$.

Although some of the phenotypes of *ssa2* mutants have been tested previously (Table 5.2), the results presented here are somewhat different compared to those reported previously. Moreover the analyses presented closely compare the phenotypes of both the *csp6-75* and *csp6-95* mutations (renamed *ssa2-75* and *ssa2-95* respectively) for the first time.

A comparative plating assay was used to compare the silencing defects, temperature sensitivity and TBZ sensitivity of the two mutants (Figure 5.1), showing that they are quite different, with *ssa2-75* displaying very mild silencing defects, no obvious TBZ sensitivity, wildtype growth at the permissive

temperature and a very low level of growth at the restrictive temperature, whereas *ssa2-95* displays severe silencing defects, marked TBZ sensitivity, reduced growth at the permissive temperature and no growth at the restrictive temperature. Comparison of the severity of phenotypes in both with those of an RNAi mutant (*dcr1*Δ) and wildtype cells allow the differences to be summarised as follows: *dcr1*Δ>*ssa2-95*>*ssa2-75*>wt.

Northern analysis of siRNA levels and pericentromere repeat transcript analysis by RT-PCR show the first discrepancy with previous results. In this study I have shown that *ssa2-75* has slightly reduced siRNA levels with a corresponding slight accumulation of transcript levels, whereas *ssa2-95* has less than a tenth of the wildtype levels of siRNA with a correspondingly strong increase in pericentromeric repeat transcript abundance. These results, together with the growth assays, suggest that both *ssa2* mutants display the classical phenotypes that are indicative of a defective RNAi pathway and confirm that the *ssa2-95* allele has stronger defects than *ssa2-75*. These results contrast with previous findings in which *ssa2-75* appeared to exhibit complete loss of both siRNA and transcripts (Portoso, 2005). However in the previous northern analysis of siRNAs even the wildtype siRNA signal was faint. Since then detection of these siRNAs has improved so it is likely that the previous analyses simply lacked the sensitivity to pick up the lower siRNA levels in *ssa2-75*. I have also found that the levels of pericentromeric repeat transcripts in the *ssa2-75* allele are much lower than the more severely defective *ssa2-95* allele, so a slight increase in transcript levels may also have been missed in these previous analyses. Further experiments on siRNA levels in the reconstructed *ssa2-75* and *ssa2-95* mutants are in the following chapter and are consistent with these analyses.

My analysis of pericentromeric H3K9me2 levels by Chromatin IP (ChIP) followed by real-time PCR (Figure 5.3) shows that *ssa2-95* retains levels of H3K9me2 that are at least as high as wildtype. Oddly the H3K9me2 levels in *ssa2-95* appear to be higher than wildtype, although this is mainly due to one in three of the replicates displaying exceptionally high enrichment of H3K9me2

whilst the other replicates were similar to wildtype. This result also differs to that of Portoso 2005 where H3K9me2 was thought to be entirely lost. However the previous analysis relied on multiplex PCR, and no PCR product was detected in the *ssa2-95* sample, even for the internal control actin. This suggests that the PCR reaction may have simply failed in this case, leading to an erroneous result.

The splicing assay (Figure 5.4) indicates that the *ssa2* mutants are not defective in splicing. As the *ssa2* mutants were later identified as Hsp70 alleles, this assay helps to rule out a possible indirect mode of action whereby the Ssa2 protein might cause a silencing defect through a hypothetical chaperone role in the splicing pathway.

ssa2^{+}, a gene encoding a Hsp70 protein, was successfully identified as the gene mutated in the *csp6* alleles through whole-genome sequencing. This method for sequencing mutations in strains has proven to be a powerful way to identify candidate genes which can then be ruled in or out by further experimental tests. Whole-genome sequencing has been shown to be a useful method for identifying mutations in fission yeast (Irvine et al., 2009). Initial problems encountered with the genome resequencing technology appear to be due to insufficient genomic coverage along with insufficient read quality. Our first strategy was to aim for a mean coverage of >8 reads per base. Although we managed to get a mean coverage ranging from 9 to 29 reads per base depending on the sample, the sequencing of the 12.8 Mb genome provided patchy, random coverage where some regions were vastly over-represented and some regions were under-represented. Concentrating on the percentage of the genome with <3 reads per base proved to be a much more reliable indicator of adequate coverage. My analysis suggests that identification of mutants requires that <0.1% of the genome should have <3 reads per base to ensure sufficient sequence accuracy. Intriguingly a superficial comparison of the low coverage (<3 reads per base) and high coverage (>twice the mean reads per base) regions for the two *ssa2* mutant genomes shows that many of the same regions display similar levels of extreme coverage. This could be attributed to the reported GC bias in the PCR step of

genomic library preparation or the GC bias caused by dense clustering of AT-rich sequences on the sequencing slide (Aird et al., 2011).

The identity of *csp6* was confirmed in two complementary approaches: the *ssa2-95* mutation was recreated in a previously wildtype strain; a *csp6-95* strain was rescued to wildtype using the *ssa2*$^{+}$ DNA (Figure 5.5). The position of the *ssa2*$^{+}$ gene on chromosome III is consistent with the previous genetic mapping (Portoso, 2005). The *ssa2*$^{+}$ gene has previously been connected to the *csp6* mutants as it consistently complimented the ts phenotype as a high copy plasmid suppressor. Isolates from genomic libraries recovered two other Hsp70 genes (*ssa1*$^{+}$ and *ssp1*$^{+}$) which also complemented the *csp6* ts phenotype (Ekwall, 1999), (Portoso, 2005). As numerous Hsp70 genes were found to rescue the temperature sensitive phenotype it was thought that this genetic complementation was due to their chaperone activity rescuing a defective Csp6 protein. Furthermore *ssa2*$^{+}$ in particular was ruled out in previous work using a genetic test which found that the *csp6* gene was not linked to the Septin gene *cdc11*$^{+}$ which is 9.7kb from the *ssa2*$^{+}$ gene (Portoso, 2005). In this assay the ts *csp6-75* allele was crossed to the *cdc11-21* ts allele with the expectation that 100% of the progeny would be ts if the alleles were linked. Only 40% of the progeny were ts however. Although the septins and Hsp70 have not been shown to have genetic interaction before it is possible that the combination of the mutants may have had an unpredicted affect. Alternatively the use of the less defective *csp6* allele in these crosses may have resulted in underscoring of *csp6-75* ts progeny in relation to *cdc11-21* progeny.

The identification of the mutations in *ssa2-75* and *ssa2-95* (Table 5.5) has shown that both mutations result in the substitution of the same amino acid. This is surprising because the alleles have very different phenotypes, with *ssa2-95* displaying much more extreme growth and silencing defects. The biochemical effects of the substitution of alanine 52 for a valine residue (*ssa2-95*) or alanine 52 for a threonine residue (*ssa2-75*) would not be expected to cause strong phenotypes because as all three amino acids are uncharged and relatively similar in size, although in both cases there is a slight increase in size. As discussed

earlier it is most likely that these mutations are affecting specific interactions with the Hsp70 cofactor proteins, J proteins and Nucleotide Exchange Factors (NEFs). These proteins act as guides for the Hsp70 peptide folding machine, providing substrate binding specificity and controlling the kinetics of substrate binding (Kampinga and Craig, 2010). The kinetics of substrate binding are important because if a protein is not bound for long enough then it may not have time to refold and if it binds for too long then the substrate and the Hsp70 are effectively sequestered (Cyr, 2008). As little analysis of the heat-shock proteins has been performed in *S. pombe* it is not known which J proteins and NEFs interact with Ssa2. This means that any follow up experiments to determine the role of specific Hsp70 cofactors in the phenotypes of *ssa2-75* and *ssa2-95* require extensive investigation of Ssa2 binding partners.

Interestingly it has recently been found that Hsp70 and Hsp90 are involved in the RNAi pathway in plants and flies (Iwasaki et al., 2010), (Iki et al., 2010), (Miyoshi et al., 2010). Specifically these studies show that the Hsp70 and Hsp90 chaperone proteins are required for efficient loading of siRNAs into Argonaute in an ATP-dependent manner *in vitro*. Thus this new link between Hsp70 and the RNAi effector protein Argonaute provides a clear direction for further investigation of the nature of the silencing defect displayed by *ssa2-75* and *ssa2-95*.

Chapter 6 - Investigating the role of the Hsp70 protein Ssa2 in RNAi-directed heterochromatin formation

6.1 Introduction

In Chapter 5 I identified the two *c*entromere: *s*uppressor of *p*osition effect (*csp*)*6* silencing defective mutants to be alleles of the Hsp70 gene *ssa2*$^{+}$. Ssa2 is one of two cytoplasmic Hsp70 proteins in fission yeast. Ssa1 and Ssa2 have >94% sequence identity at the amino acid level.

Hsp70 proteins are chaperones which fold or refold proteins in an ATP-dependent manner and this function is particularly important in the cell's response to high temperatures for protecting proteins from denaturation. The Hsp70 family of proteins is very highly conserved across eukaryotes and numerous different Hsp70 proteins are present in all species. Numerous cofactor proteins, known broadly as Nucleotide Exchange Factors (NEFs) and the J Domain family of proteins, interact with Hsp70 proteins in order to regulate Hsp70 function. These cofactors interact with Hsp70 via its Nucleotide Binding Domain (NBD) and can affect the speed of ATP/ADP turnover which in turn affects substrate binding and folding kinetics (Kampinga and Craig, 2010) (Figure 6.1). Hsp70 in an ATP-bound conformation undergoes a fast capture and release of substrate proteins, whereas the ADP-bound confirmation holds the substrate protein more stably. J domain proteins can bind unfolded proteins to recruit them to Hsp70, and then increase the efficiency of Hsp70 ATPase activity in order for Hsp70 to stably bind the substrate/unfolded protein (Szabo et al., 1994). NEFs consequently allow the release of substrate proteins from Hsp70 through nucleotide exchange, returning the Hsp70 from an ADP-bound state to an ATP-bound state (Szabo et al., 1994).

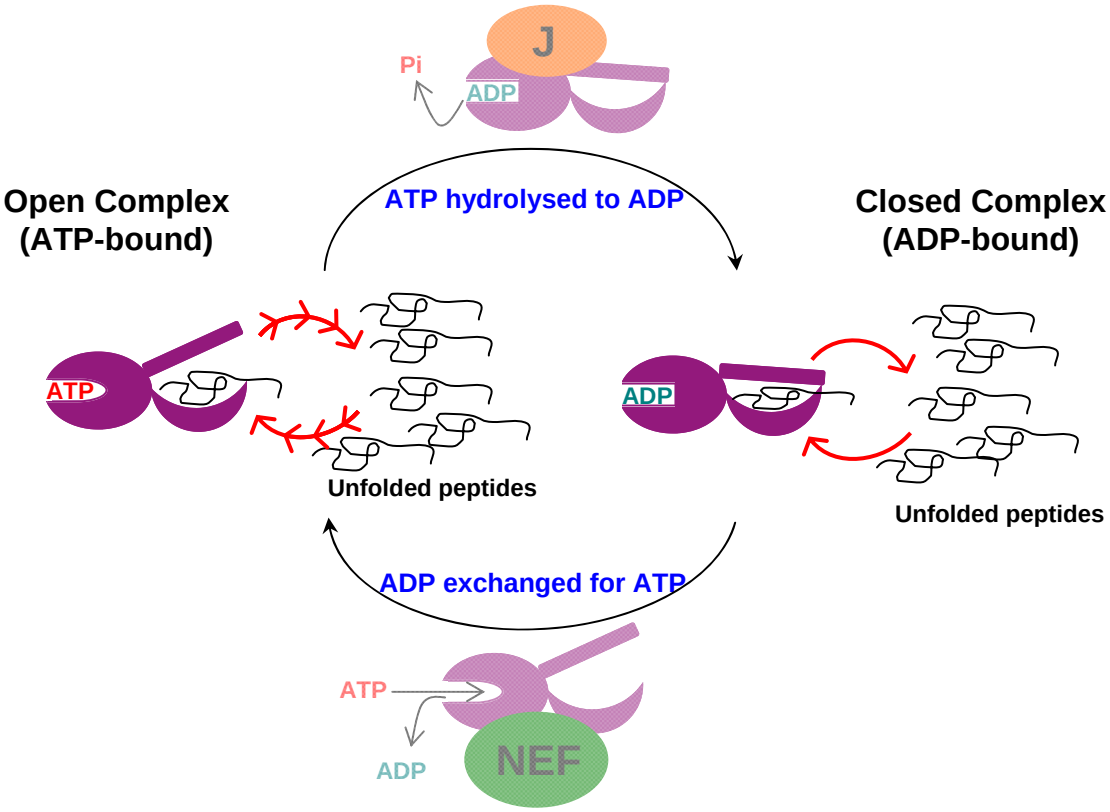


Figure 6.1 – The nucleotide binding cycle of Hsp70.
The nucleotide bound in the nucleotide binding domain (NBD) of Hsp70 affects the conformation of the protein, thus allowing regulation of its activity. When ATP is bound to the NBD then the substrate binding domain (SBD) is in an open conformation and there is fast binding and release of substrate peptides. When ADP is bound at the NBD then the SBD is in a closed conformation where the substrate is stably captured by the chaperone.
The correct balance between substrate binding/release and stable substrate capture is required for optimum substrate (re)folding thus correct regulation of the ATP and ADP bound states of Hsp70 is vital.
Substrate binding and J proteins stimulate the ATPase activity of the NBD causing hydrolysis of ATP to ADP. Nucleotide exchange factors (NEFs) stimulate release of ADP from the NBD allowing consequent binding of ATP.

Recently it has been shown that the heat-shock protein (Hsp)90 (Iwasaki et al., 2010), (Iki et al., 2010), (Miyoshi et al., 2010) and Hsp70 proteins (Iwasaki et al., 2010), (Iki et al., 2010) are required for loading of siRNAs into Argonaute. These studies used *in vitro* systems to test the effects of Hsp90 and Hsp70 inhibitors (geldanamycin and 2-phenylethynesulfonamide respectively) on the assembly and activity of the RNA-induced Silencing Complex (RISC) using

plant or fly cell extracts. These inhibitors were found to dramatically reduce the loading of double stranded siRNA into the Argonaute protein. The loading of siRNAs was also found to be ATP-dependent, which is assumed to reflect the ATPase activity of these heat-shock proteins.

In this chapter I investigate the cause of the silencing phenotype of the *ssa2-75* and *ssa2-95* alleles which affect the function of a Hsp70 protein. I present experiments to test whether this Hsp70 plays a role in loading siRNAs into Ago1 (the only Argonaute protein in *S. pombe*) and if this is actually responsible for the silencing defects seen in the *ssa2* mutants.

6.2 Results

6.2.1 The *ssa2-75* and *ssa2-95* mutants have a competitive inhibitory effect on pericentromeric silencing

It is known that neither of the cytoplasmic Hsp70 proteins of *S. pombe* are essential for cell viability (Kim et al., 2010). In order to test the phenotype associated with deletion of the *ssa2*$^+$ gene on heterochromatin function I replaced the *ssa2*$^+$ open reading frame with the cloNAT antibiotic resistance gene. This *ssa2*Δ was generated in the same genetic background used to generate the original *ssa2* point mutants, *ssa2-75* and *ssa2-95*, with the *ade6*$^+$ reporter gene in the pericentromeric repeats of cen1. A serial dilution comparative plating assay shows that, in contrast to the *ssa2-75* and *ssa2-95* mutants, *ssa2Δ* does not exhibit defective silencing (Figure 6.2). This assay demonstrates that the *ssa2*$^+$ gene is not essential for pericentromeric chromatin integrity. This is surprising since the *ssa2-75* and *ssa2-95* mutants clearly affect silencing at pericentromeric repeats and suggest that the *Ssa2-75 and Ssa2-95* mutant proteins are causing a defect in silencing because of their disruptive presence in the RNAi/heterochromatin system, presumably through a change in the way the Hsp70 protein is functioning. A possible explanation for this 'competitive inhibitory' effect of the *Ssa2* point mutant proteins is that Ssa1 and Ssa2 may act redundantly in RNAi-

dependent silencing; mutation of Ssa2 may both inhibit its own function, and, due to competition between Ssa1 and Ssa2 for binding sites, also block the function of Ssa1.

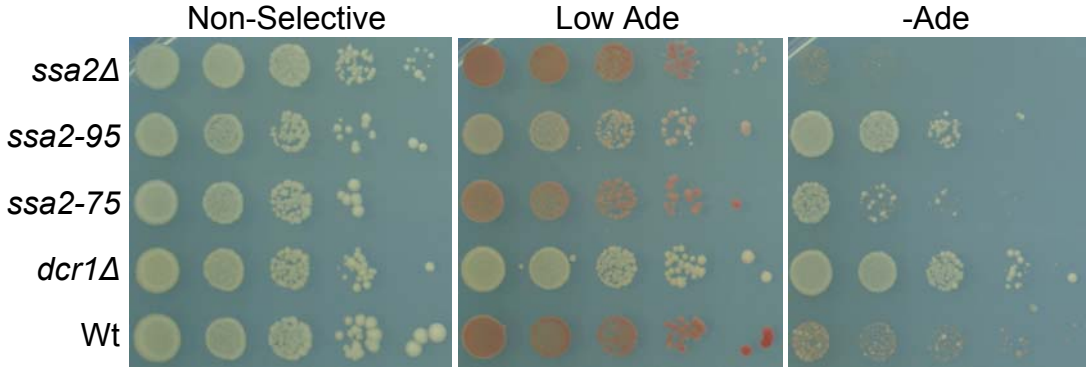


Figure 6.2 – The *ssa2* point mutants display a competitive inhibitory effect on silencing. A serial dilution comparative plating assay was used to test heterochromatic silencing at the pericentromere in an *ssa2Δ* mutant. The ade6 gene, which is situated in the pericentromeric heterochromatin region, allows the testing of silencing through growth on low adenine plates (wt is red, defective silencing is white) and minus adenine plates (no growth for wildtype, good growth for defective silencing).

6.2.2 Ssa2 associates with Ago1

The recent finding regarding the role of Hsp70 and Hsp90 in loading Argonaute with siRNA showed that chaperone proteins interact with Argonaute proteins and affect siRNA loading *in vitro* in plant or fly extracts (Iwasaki et al., 2010), (Iki et al., 2010), (Miyoshi et al., 2010). To determine if Ago1 (the only Argonaute protein in *S. pombe*) interacts with Ssa2 *in vivo* in *S. pombe*, a Co-immunoprecipitation (CoIP) was performed. Both the wildtype $ssa2^{+}$ and mutant *ssa2-95* genes were tagged at the C-terminus with 3xHA and tested for association with Ago1. Flag-tagged Ago1, Clr4 and Rdp1 were combined with *ssa2-HA* and *ssa2-95-HA* by crossing. Analysis of immunoprecipitates by western analysis shows that both the mutant and wildtype Ssa2 proteins interact with Ago1 to the same degree (Figure 6.3). Ssa2 is also pulled down by other proteins involved in RNAi-dependant heterochromatin formation; the H3K9 methylating

enzyme Clr4 and the RNA-dependent RNA Polymerase Rdp1. These proteins are part of different complexes in RNAi-dependent heterochromatin formation (Figure 6.4), but it is known that the interaction of both of these complexes with the Ago1-containing RITS complex can be detected (Motamedi et al., 2004), (Zhang et al., 2008). It is therefore possible that Ssa2 only directly binds to one of the proteins tested and that the other interactions detected are indirect, through the interaction between the different complexes involved in RNAi-dependent heterochromatin formation.

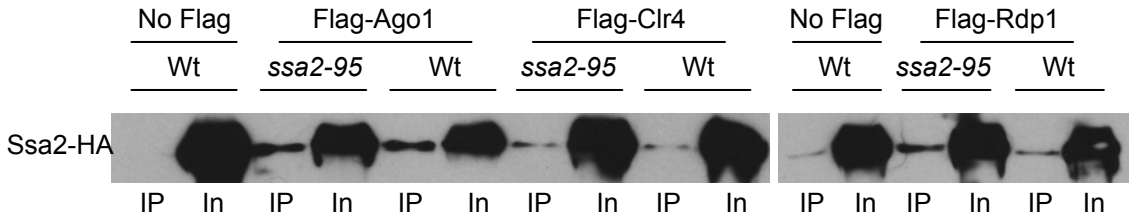


Figure 6.3 – The *ssa2-95* mutation does not affect the association of Ssa2 and proteins involved in RNAi-mediated heterochromatin formation. Initially the Flag-tagged proteins were immunoprecipitated with α-Flag antibody. The pulldowns were then separated by size by SDS-PAGE and the western blot was probed using α-HA antibody to detect the HA-tagged Ssa2 protein. Cultures were grown at the permissive temperature (25°C).

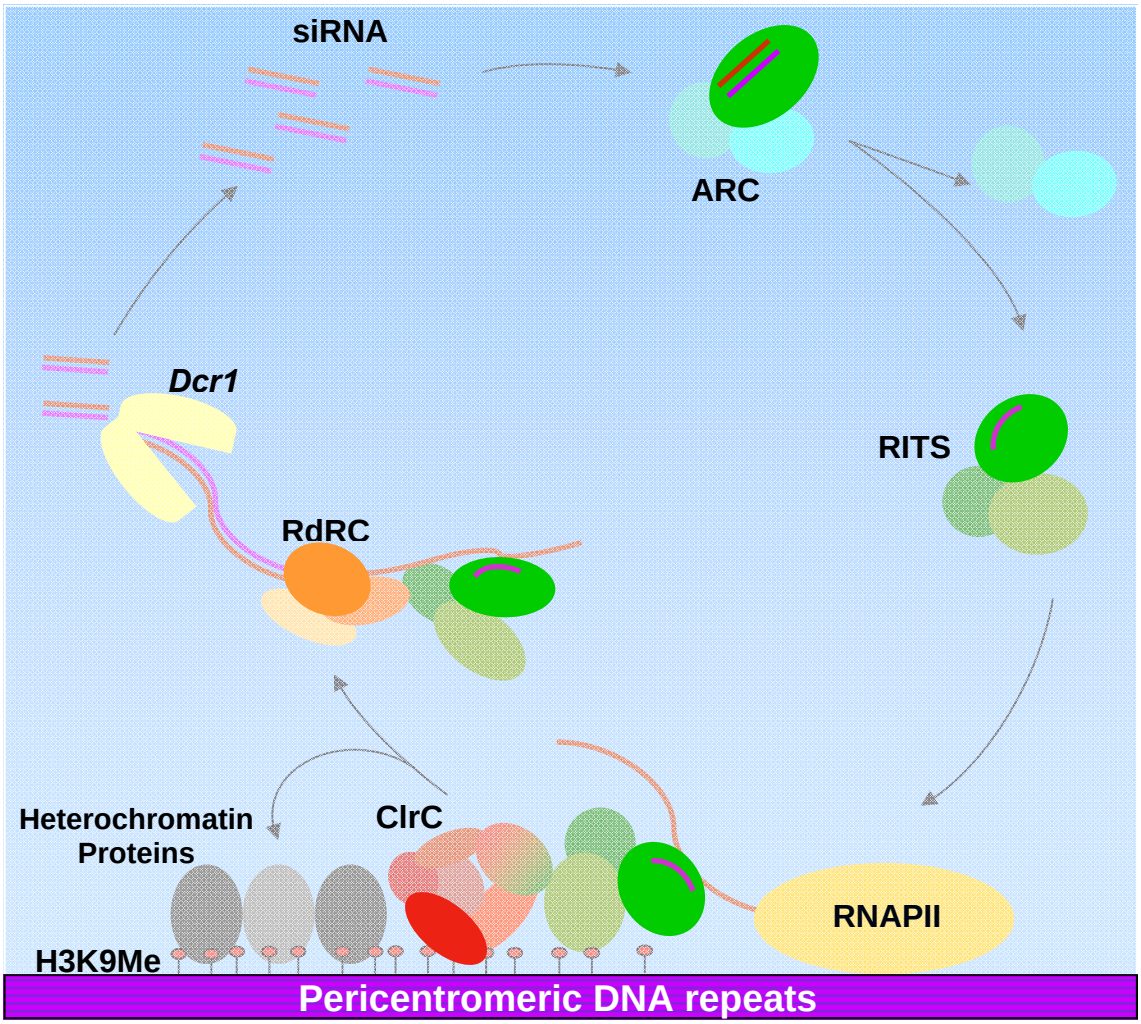


Figure 6.4 – RNAi-mediated heterochromatin formation in *S. pombe*. The highlighted proteins Ago1, Rdp1 and Clr4 were Flag-tagged and immunoprecipitated in order to test their association with the HA-tagged Ssa2 protein.

6.2.3 Assaying the levels of siRNA associated with Ago1 in *ssa2-75* and *ssa2-95*

In order to investigate if siRNAs are being loaded into Ago1 in the cells bearing *ssa2-75* and *ssa2-95*, flag-tagged Ago1 was immunoprecipitated and the Ago1-associated siRNA levels were assessed by northern analysis (Figure 6.5).

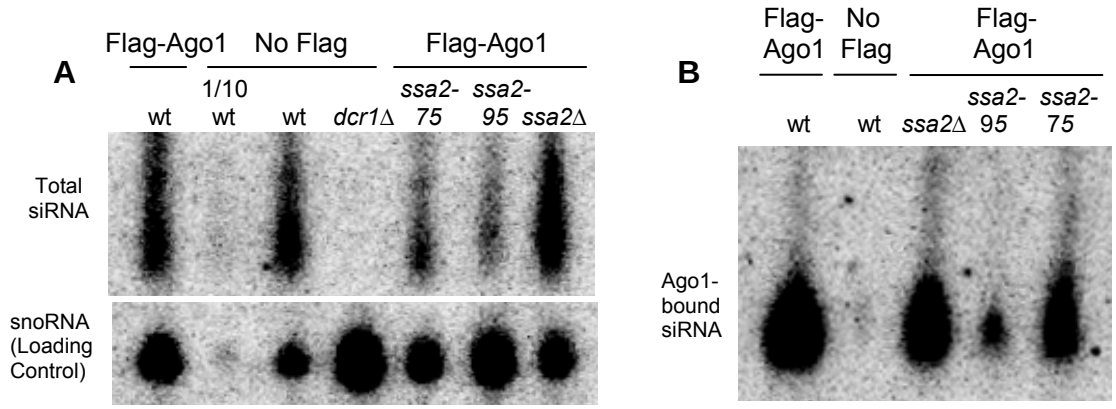


Figure 6.5 – Comparison of whole-cell siRNA levels with Ago1-associated siRNA levels using northern analysis (A) Total levels of siRNA in the whole cell. (B) Ago1-associated levels of siRNA determined by IP pull-down of Flag-tagged Ago1 followed by siRNA extraction and northern analysis.

This analysis demonstrates that siRNAs are still loaded into Ago1 in both *ssa2-75* and *ssa2-95* mutants. Less siRNAs are associated with Ago1 in *ssa2-95* cells than in *ssa2-75*, *ssa2*Δ or wildtype cells. However, comparison of the whole-cell and Ago1-associated siRNA levels suggests that the amount of siRNA present in Ago1 is approximately proportional to the amount of total siRNA in the cell. This data could be explained in several ways: (1) Ago1 loading is less efficient in both *ssa2-75* and *ssa2-95* cells so less siRNA is loaded and the increased free siRNA is degraded, leading to a reduction in the observed total siRNA levels (2) Ago1 loading is less efficient in *ssa2-75* and *ssa2-95* so less siRNA is loaded therefore less of the Ago1-containing RITS complex is targeted to homologous transcripts which results in a reduction of siRNA being made (3) The efficiency of Ago1 loading is unaffected but the available levels of total siRNA are reduced through another mechanism that affects RNAi-dependent heterochromatin

formation (4) The efficiency of loading Ago1 with siRNA is unaffected but after loading the defective Ssa2 protein does not release Ago1 and consequently disrupts the RNAi-directed heterochromatin formation pathway.

Eri1 is a ribonuclease which has been shown to degrade double-stranded siRNA (Iida et al., 2006). Ago1-associated siRNA was compared to total siRNA levels in an *eri1Δ* background in order to assess if the low levels of total siRNA in *ssa2-95* are due to Eri1-mediated degradation of unloaded siRNAs. Again northern analysis was used to compare total levels of siRNAs with levels of Ago1-associated siRNAs (Figure 6.6). A noticeable increase in both total and Ago1-associated siRNA was observed in *ssa2-95* cells from which *eri1*$^{+}$ had been deleted, suggesting that the efficiency of siRNA loading is not affected by *ssa2-95.* Also the fact that there is a similar increase in siRNA associated with Ago1 in wildtype and *ssa2-95* supports this conclusion.

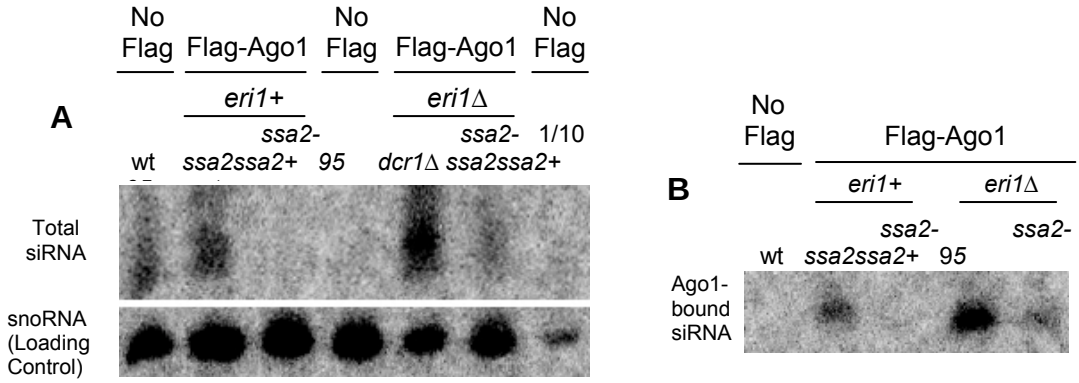


Figure 6.6 – siRNA levels in an *eri1Δ* and/or *ssa2-95* mutant background compared against wildtype. Eri1 is required for degradation of unloaded siRNA thus siRNA levels are higher in an *eri1Δ* mutant (Iida 2006). Whole-cell siRNA levels and Ago1-associated siRNA levels are detected using northern analysis. (A) Total levels of siRNA in the whole cell. (B) Ago1-associated levels of siRNA determined by IP pull-down of flag-tagged Ago1 followed by siRNA extraction and northern analysis.

To test the hypothesis that the mutated Ssa2 is not releasing Ago1 after binding, the promoter of the *ago1*$^{+}$ gene was replaced with the inducible *nmt1* promoter which will over-express Ago1 after induction in medium lacking

thiamine. Over-expression of Ago1 may be expected to rescue the silencing defect seen in *ssa2-75* and *ssa2-95* if it is caused by the sequestering of Ago1. However a serial dilution comparative plating assay of these strains shows that the overexpression of Ago1 does not rescue the silencing defects seen in the *ssa2-75* and *ssa2-95* cells (Figure 6.7). Thus I am unable to conclude that defective silencing in the *ssa2* mutants results from sequestration of Ago1.

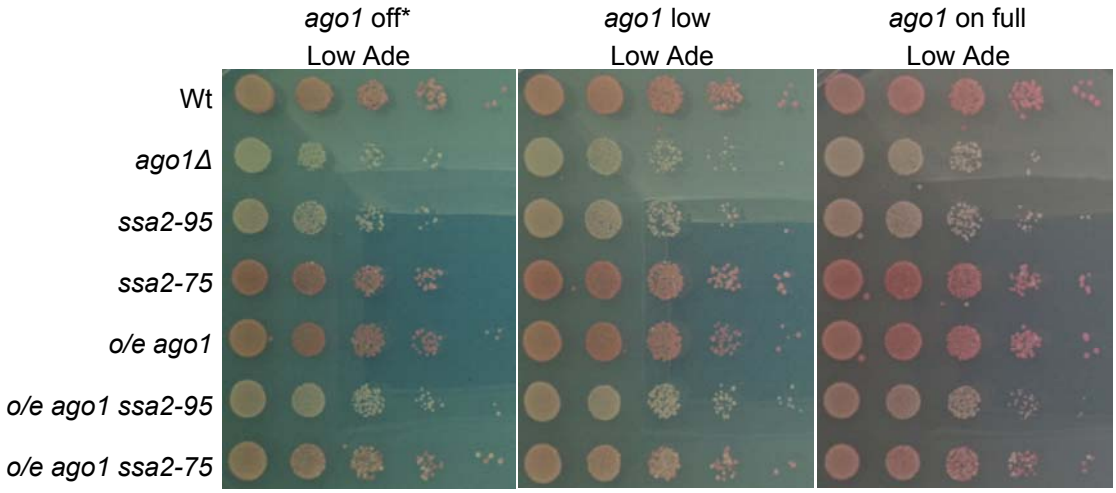


Figure 6.7 – Overexpression of Ago1 does not rescue the silencing defects of *ssa2-75* and *ssa2-95*. A serial dilution comparative plating assay was used to test the effects of Ago1 overexpression (o/e) on the *ssa2-75* and *ssa2-95* mutants. All strains harbour *ade6+* positioned in the pericentromeric heterochromatin. *ade6+* silencing is tested using growth on low adenine plates where wildtype silencing causes a red colour and defective silencing causes a white colour. *o/e ago1*: the endogenous promoter for the *ago1+* gene was replaced by the *nmt1* promoter which has a very high expression level in cells grown in a lack of thiamine. Plates contain 15μM thiamine (ago1 off), 0.06μM thiamine (ago1 low) or no thiamine (ago1 on full).

6.3 Discussion

In this chapter I investigated the role of Ssa2, a cytoplasmic heat shock protein (Hsp)70, in the RNAi pathway. I have shown that although *ssa2-75* and *ssa2-95* cells have a silencing defect, however deletion of the *ssa2*$^{+}$ gene does not result in a defect in pericentromeric heterochromatin formation (Figure 6.2). This result indicates that the *ssa2-75* and *ssa2-95* mutants in *ssa2*$^{+}$ have a competitive inhibitory effect on RNAi-directed heterochromatin formation. To determine which components of RNAi-directed heterochromatin formation the Ssa2 protein interacts with to cause the RNAi defects, co-immunoprecipitation analyses were carried out using Flag-tagged versions of several RNAi and heterochromatin proteins to test for association with Ssa2-HA, in wildtype and *ssa2-95* cells. Disappointingly, the Ssa2 wildtype and mutant proteins were found at the same levels in all immunoprecipitates tested (Figure 6.3). This suggests that these interactions are non-specific and that a specific interaction with a particular component may yet underlie the role of Ssa2 in mediating effective RNAi.

The promiscuous binding of Ssa2 detected in immunoprecipitation experiments makes it difficult to determine whether an RNAi protein might be affected by *ssa2-75*/*ssa2-95* to cause the observed defects in silencing. In addition to the RNAi proteins tested for their interaction with Ssa2, the unrelated protein Scm3 was also found to interact with Ssa2 (Appendix - Figure 6) confirming the ubiquitous role of Ssa2. The promiscuity of Ssa2 protein binding is a well known phenomenon, as Hsp70 proteins are notorious contaminants in large scale affinity purification experiments (Gavin et al., 2002), (Roguev et al., 2004). As well as numerous possible targets in the RNAi pathway, there are also numerous ways in which the *ssa2-95* mutation may be causing a defect, for example: (1) Inefficient siRNA loading into Ago1 (2) Inefficient capture of an RNAi protein or complex (3) Inefficient release of an RNAi protein or complex.

Hsp70 proteins affect the loading of Argonaute with siRNAs (Iwasaki et al., 2010), (Iki et al., 2010) the levels of Ago1-associated siRNA was tested in

ssa2 mutants by comparing the total levels of siRNA in the cell with the amount of siRNA associated with Ago1. (Figure 6.5). Although the level of siRNAs loaded into Ago1 in *ssa2-75* and *ssa2-95* cells was less than that observed in wildtype cells, the total siRNA levels were essentially proportional to the levels of siRNA associated with Ago1. To test if the reduced levels of total siRNA in *ssa2-95* are due to inefficient loading and consequent degradation of siRNA, the level of total siRNAs was compared with Ago1-Flag associated siRNAs in an *eri1*Δ background in which unloaded siRNA is not degraded (Figure 6.6). An increase in both total and loaded siRNAs was observed in *ssa2-95* suggesting that the efficiency of siRNA loading is unaffected by *ssa2-95.* Together these analyses suggest that silencing defects are likely caused by another mechanism. A possible flaw in this argument is that the degradation of siRNA has not been thoroughly investigated in fission yeast thus although elevated levels of siRNA are detected in *eri1*Δ cells it is not known if Eri1 is the nuclease involved in degrading free siRNAs.

One problem with testing the siRNA loading by comparative analysis of total siRNA levels with Ago1-associated siRNA levels is that the effects of inefficient siRNA loading cannot be fully distinguished from a general defect in the RNAi pathway. In order to address this, a GFP hairpin system could be used (Sigova et al., 2004), (Simmer et al., 2010). The GFP hairpin produces dicer-dependent siRNAs, the levels of which are produced independent from downstream events within the RNAi pathway and thus measuring loading efficiency of the siRNA into Ago1. However, the GFP hairpin system has numerous practical drawbacks, mainly due to the fact that the GFP hairpin is very highly expressed from the *nmt1* promoter which is only expressed in minimal media lacking thiamine. Normally 2L of cells are grown in 4xYES to a concentration of $2x10^8$ cells/ml and used to analyse Ago1-associated siRNAs. In minimal media (PMG) cells can only be grown to a concentration of $1x10^7$ cells/ml thus 40L of culture would be required per sample. Due to the high expression level of the GFP hairpin RNA the siRNA produced are partially masked by the products of GFP RNA degradation (Simmer et al., 2010). The

masking of the siRNA is the main problem as comparison of the total siRNA levels to the Ago1-associated siRNA levels is vital. Although this can be partly addressed by including a *dcr1*Δ control for GFP degradation background levels, our experience suggests that such analyses will be inconclusive.

The fact that the *ssa2*Δ mutant has no silencing defect (Figure 6.2) suggests that although the *ssa2-75* and *ssa2-95* mutants are interfering with RNAi, Ssa2 does not normally play an essential role in the RNAi pathway. This argues against the possibility that reduced efficiency of capture of RNAi proteins by Ssa2 is the cause of the silencing defects observed in the *ssa2-75* and *ssa2-95* mutants. A more likely scenario is that Ssa2 normally functions redundantly with Ssa1, but Ssa1 function is also blocked in *ssa2-75* and *ssa2-95* cells due to a dominant affect of the mutant Ssa2. An interesting experiment to test this redundancy would be to make the identical amino acid substitutions in the Ssa1 protein to determine if *ssa1* mutants also cause a silencing defect. Alternatively, silencing could be tested in an *ssa1*Δ*ssa2*Δ double mutant to test the redundancy in the RNAi-dependent heterochromatin formation pathway.

The efficiency of release of RNAi proteins such as Ago1 by Ssa2 is implicated as the most likely mechanism for the silencing defects in *ssa2-75* and *ssa2-95* cells because these mutants have a competitive inhibitory effect on silencing. If *Ssa2-95* effectively sequesters proteins, then it would be predicted that this more stable binding might be detectable as an increased association in immunoprecipitation. However the analyses do not show a difference in binding of *Ssa2-95* to Ago1, Rdp1 or Clr4 compared to wildtype. In addition, overexpression of the Ago1 protein failed to rescue the silencing defects of *ssa2-95* and *ssa2-75* although this could be because Ago1 is not the target protein. Also a western blot is required to confirm that Ago1 is over-expressed when the *ago1*$^{+}$ gene is expressed under the control of the *nmt1* promoter. Thus it remains debatable that inefficient release of RNAi client proteins causes the defective heterochromatin formation phenotypes in *ssa2-75* and *ssa2-95*.

A disadvantage of the overexpression experiment used to test the client protein release hypothesis is that the actual RNAi target protein is unknown and

testing each potential target would require the same replacement of its promoter. During previous attempts to identify the *ssa2-75* and *ssa2-95* mutations, high copy plasmid borne genomic libraries which moderately over-express all genes, were used to rescue the temperature sensitive phenotypes of *ssa2-75* and *ssa2-95*. Similar strategies could be employed to identify the critical client protein by looking for the gene(s) that rescue the silencing defects of *ssa2-95* mutant. However, the variegating phenotype (ie the appearance of both red and white colonies due to spreading/receding of heterochromatin over the *ade6*$^{+}$ marker gene) observed in *ssa2-95* (Figure 6.2) makes this selection strategy impossible because it leads to a high rate of false positives which negates the use of this approach. As similar variegation phenotypes have been observed when using the *ura4*$^{+}$ reporter gene (Appendix - Figure 3) the variegation problem is not specific to *ade6*$^{+}$ expression.

Clearly additional detailed analyses are required to determine the mechanism underlying the silencing defect in the *ssa2-95* mutant. However, the simplest conclusion is that the analysis presented here does not support a role for Hsp70 in loading siRNA into Ago1 in *S. pombe in vivo*. This suggests that the role of Hsp70 in loading plant and *Drosophila* Argonaute with siRNA (Iwasaki et al., 2010), (Iki et al., 2010) may not be conserved. However the original observations used an *in vitro* assay thus a similar *in vitro* assay would need to be implemented to test if Ssa2 equivocally does or does not play a role in siRNA loading of Ago1.

Chapter 7 – Discussion and conclusions

The centromere is a critical region of the chromosome as it provides the platform on which the kinetochore assembles (Karpen and Allshire, 1997). Kinetochore assembly is essential for regulated microtubule attachment to the chromosome in order for equal segregation of DNA into the daughter cells during cell division. Fission yeast centromeres are similar to metazoan centromeres in that regions of centromeric chromatin containing the H3-variant CENP-A are situated adjacent to regions of heterochromatin. The centromeres of mammals are made up of highly repetitive regions of DNA (alpha-satellite DNA) which are assembled into alternating regions of centromeric chromatin and heterochromatin (Blower et al., 2002). This is distinct from the centromeres of fission yeast which contain a central unique sequence packaged in centromeric chromatin which is flanked on both sides by the outer repeats which are heterochromatic (Partridge et al., 2000).

The heterochromatin at the pericentromeric outer repeats of *S. pombe* is formed by the RNAi-dependent heterochromatin formation pathway (Volpe et al., 2002). Cells lacking RNAi proteins retain healthy growth, but they display lagging chromosomes due to the role of pericentromeric heterochromatin in chromatid cohesion during cell division (Volpe, 2003). Currently we have a good understanding of the role of many of the proteins encoded by genes not essential for cell viability, such as components of the RITS, ClrC, RdRC complexes in this pathway. Some essential proteins have also been found to play a part in the pathway such as splicing factors (Bayne et al., 2008) and RNAPII (Djupedal, 2005), (Kato, 2005) however the roles of these proteins are poorly understood. In this thesis I have studied the role of proteins which are both essential for efficient, healthy cell growth as well as for RNAi-directed heterochromatin formation. I have described the generation and analysis of a number of RNAPII mutants which affect RNAi-dependent heterochromatin formation. I have also identified the gene affected in the *csp6* ts mutants to be $ssa2^{+}$ which encodes a Hsp70 protein.

7.1 The role(s) of RNAPII in RNAi-dependent heterochromatin formation

RNAPII is the RNA polymerase responsible for transcribing the bulk of ncRNAs in metazoans (Seidl et al., 2006) although the plant-specific RNAPIV is responsible for transcribing ncRNAs in plants (Onodera et al., 2005), (Herr et al., 2005). In fission yeast RNAPII transcribes the outer repeats of the pericentromere (Djupedal, 2005) and the transcript is processed into siRNAs by the RNAi machinery. Transcripts originating from pericentromeric repeat sequences have also been detected in many organisms such as mice (Rudert et al., 1995), (Martens et al., 2005), humans (Valgardsdottir et al., 2005) and plants (May et al., 2005). RNAPII transcription of pericentromeric repeats is under cell cycle control in mice and fission yeast (Lu and Gilbert, 2007), (Kloc et al., 2008), (Chen et al., 2008) and in fission yeast the peak of transcription coincides with DNA replication during S phase, providing an increased pool of siRNA at a time when heterochromatin is being reinstated after replication. A recent study in *S. pombe* has also shown that two ClrC components required for silencing are present in another complex with the DNA polymerase catalytic subunit Cdc20 (Li et al., 2011). This complex was found to regulate RNAPII activity, heterochromatin formation and DNA replication.

Previous work with RNAPII mutants in *S. pombe* has shown that there are at least two roles of RNAPII in RNAi-directed heterochromatin formation: (1) transcription of ncRNA for siRNA generation (Djupedal, 2005) (2) processing of ncRNA into siRNA (Kato, 2005). As discussed in Chapter 4, I have isolated a new class of RNAPII mutant which has a strong negative effect on RNAi-dependent heterochromatin formation but a small positive RNAi-independent affect on heterochromatin formation. The roles of RNAPII in this pathway are thus very complex.

A number of RNAPII subunits have been found to physically interact with the RITS complex protein Chp1 through immunoprecipitation experiments in *S. pombe* (unpublished observation A. Kagansky, Allshire Lab). Similarly, the largest subunit of RNAPIV has been shown to physically associate with AGO4

in plants (Li, 2006). Moreover, in *Drosophila,* Rpb1 has been found to interact with Dcr-2 (which is involved in siRNA processing) and dAGO1 (which has slicer activity and associates with miRNA) through co-immunoprecipitation (Kavi and Birchler, 2009). Interaction between RNAPII and the RNAi machinery is therefore likely conserved. The physical interaction between RNAPII and RNAi components may account for one of the roles of RNAPII in RNAi-dependent heterochromatin formation whereby the RNAPII complex might act as a platform for co-transcriptional processing of pericentromeric repeat transcript into siRNA. As splicing is a co-transcriptional process and numerous splicing factors have been found to affect RNAi-dependent heterochromatin formation (Bayne et al., 2008), it is possible that splicing factors may also form part of this platform for RNAi. This role for RNAPII could account for the siRNA processing defects seen in *rpb2-m203* (Kato, 2005).

Deletion of the mRNA export factor gene *mlo3*$^{+}$ or the transcription elongation factor *tfs1*$^{+TFIIS}$ causes RNAi-independent heterochromatin formation (unpublished observation Grewal Lab), and I have observed similar effects with the RNAPII mutants examined in Chapter 4. The most likely basis of this effect is through aberrant transcription. As Tfs1 is required for rescuing stalled RNAPII (Kettenberger et al., 2003), (Kettenberger et al., 2004) it is possible that the phenomenon observed in *tfs1* mutant cells is brought about by stalled elongation or by defective release of the pericentromeric transcript by RNAPII, which promotes the recruitment of ClrC. Further investigation of this phenomenon is required, but it may indicate that RNAi normally recruits ClrC through stalling RNAPII (Figure 7.1) or preventing the release of repeat transcripts from RNAPII. Perhaps regions of heterochromatin are preferentially assembled on repetitive sequences because these regions are prone to transcriptional blocks and thus provide conditions which promote heterochromatin formation.

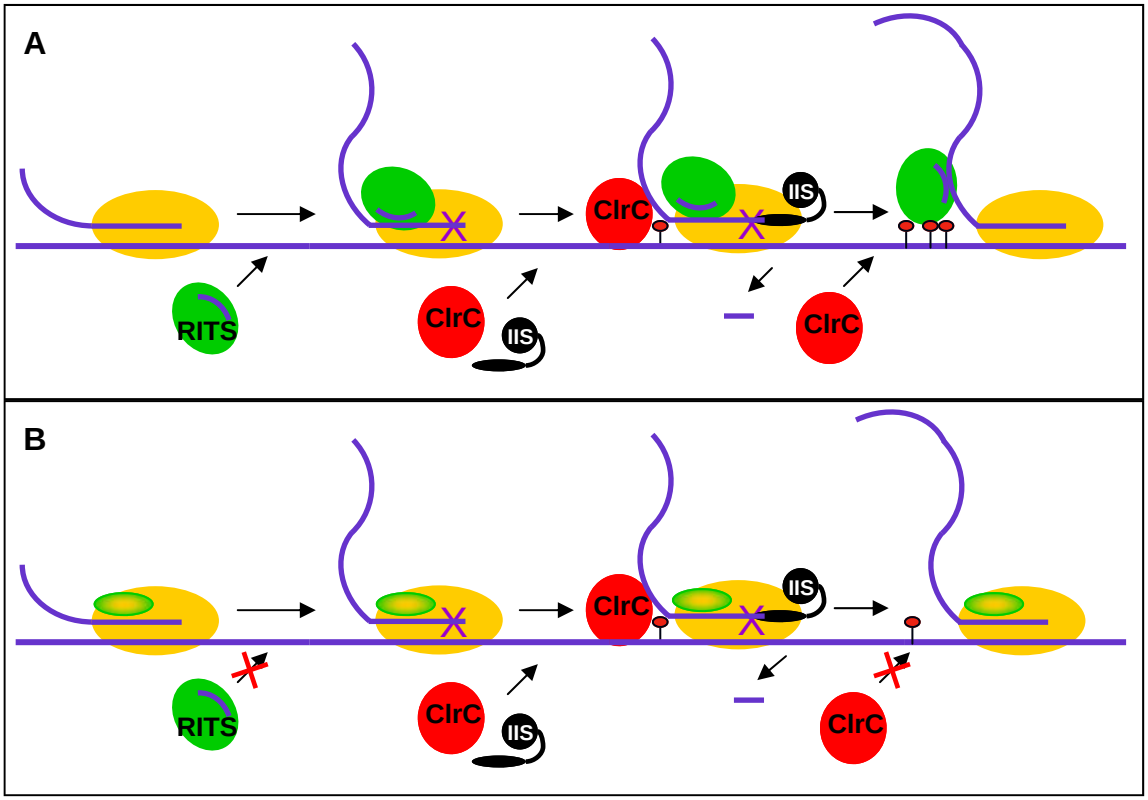


Figure 7.1 – Model of the possible RNAPII stalling-dependent recruitment of the Clr4-containing complex (ClrC).
(A) In a wildtype cell the RITS complex, guided by the Ago1-associated siRNA, may interact with the elongating RNAPII at the pericentromeric repeats and cause RNAPII stalling. The stalled RNAPII could recruit ClrC which methylates lysine 9 of the histone H3 tails nearby. TFIIS^{tfs1} can rescue the stalled RNAPII and the RITS complex can bind to the H3K9 methylated histone tails in order to recruit further ClrC for robust heterochromatin formation.
(B) Mutations in the *rpb3* and *rpb11* subunits of RNAPII may no longer allow RITS interaction with elongating RNAPII but the mutations could mimic RITS association to cause RNAPII stalling. The stalled RNAPII could recruit ClrC which methylates lysine 9 of the histone H3 tails nearby. TFIIS^{tfs1} can rescue the stalled RNAPII but without stabilising interactions between the RITS complex, H3K9 methylation and ClrC further H3K9 methylation is not produced resulting in only partial heterochromatin formation.
Purple X depicts paused RNAPII at the active site. The red lollipop depicts H3K9 methylation.
Adapted from (Palangat et al., 2005) Figure 4B

7.2 Identifying causative point mutations using genome resequencing

This technique has quickly become an important tool for identification of genes mutated during forward screens in organisms with relatively small genomes. With the 3rd generation of genome sequencing platforms coming in the near future it is likely that genome resequencing will be routinely used for analysis of

mutations in even the largest genomes, such as mammals, due to increased speed and reduced costs.

One of the main problems associated with genome resequencing is finding a balance between cost and sequencing depth. Sarin et al 2008 suggest that 8 fold coverage of the genome should be enough to produce a list of candidate genes for testing in most organisms whilst Irvine et al 2009 suggest 20-25 fold coverage. I found that the ~12 fold coverage provided for three of the *csp* mutants sequenced still left ~8% of the genome with <5 reads per base (see Chapter 5 Table 5.3). Also, even with 28 fold coverage of the genome, the *csp6-95* mutation was not identified when using the high stringency data filters, although the mutation was identified, along with 16 other SNPs, using the reduced stringency filters. This is because a combination of low coverage and a miscalled read meant that the base was not mutated in >90% of the reads. I suggest that average coverage of the genome is not the most reliable way of judging adequate genome coverage but that the percentage of the genome with low coverage is a more effective indicator.

7.3 Hsp70 mutants affect RNAi-dependent heterochromatin formation

The heat shock proteins Hsp70 and Hsp90 are encoded by multiple genes in every organism (Tavaria et al., 1996) and they are often described as house keeping genes due to their constitutively high expression levels. As described in Chapter 1.3.5.1, Hsp70 and Hsp90 can work together for processes such as regulating steroid hormone receptors (Daniel et al., 2007) or individually in a variety of different cell functions.

In vitro experiments using human proteins have shown that the interaction between the RNAi proteins Argonaute and Dicer is dependent on Hsp90 activity (Tahbaz et al., 2004). In human cells Hsp90 inhibition by geldanamycin causes decreased levels of Argonaute proteins, and it was proposed that this is due to instability of Argonaute proteins that were not yet associated with small regulatory RNAs (Johnston et al., 2010). Hsp70 and Hsp90 have been implicated

in a collaborative role in RNAi through in vitro experiments using fly or plant extracts whereby the heat shock proteins increase efficiency of loading siRNA into Argonaute (Iki et al., 2010), (Iwasaki et al., 2010).

The identification of the *csp6* mutants as alleles of the Hsp70-encoding gene *ssa2*$^{+}$, described in Chapter 5, suggests that Hsp70 proteins may also have a role in RNAi-dependent heterochromatin formation in *S. pombe*. Further experiments to detect the levels of siRNA associated with Ago1 in the *ssa2* mutant cells (Chapter 6) suggest that Ssa2 is not required for efficient loading of siRNAs in fission yeast. This conclusion is also strengthened by the observation that, unlike the *csp6* alleles, the *ssa2*Δ mutant does not have a silencing defect. This suggests that either the Ssa2 protein is not required for silencing, or else it plays a redundant role with the other cytoplasmic Hsp70 protein, Ssa1. The Hsp70 proteins may be required to promote the slicing activity of Ago1 or to increase the efficiency of siRNA passenger strand release from Ago1. Alternatively the Hsp70 proteins could be required for the stability of any of the RNAi factors or in assembly of one of the RNAi complexes. The multifunctional roles of Hsp70 proteins means that they may even be affecting RNAi-dependent heterochromatin formation at multiple stages.

The residue which is mutated in the *csp6* mutants is located in the nucleotide binding domain (NBD) of Ssa2 and thus may affect interaction of Ssa2 with its cofactors. The loss of interaction between Ssa2 and a nucleotide exchange factor (NEF) could prevent recycling of Ssa2 from an ADP-bound state to an ATP-bound state resulting in the substrate protein becoming sequestered by Ssa2 (Figure 7.2). It may be possible to identify the Hsp70 cofactors specifically involved in RNAi-dependent heterochromatin formation through overexpression or deletion of the genes previously identified as homologues from other organisms.

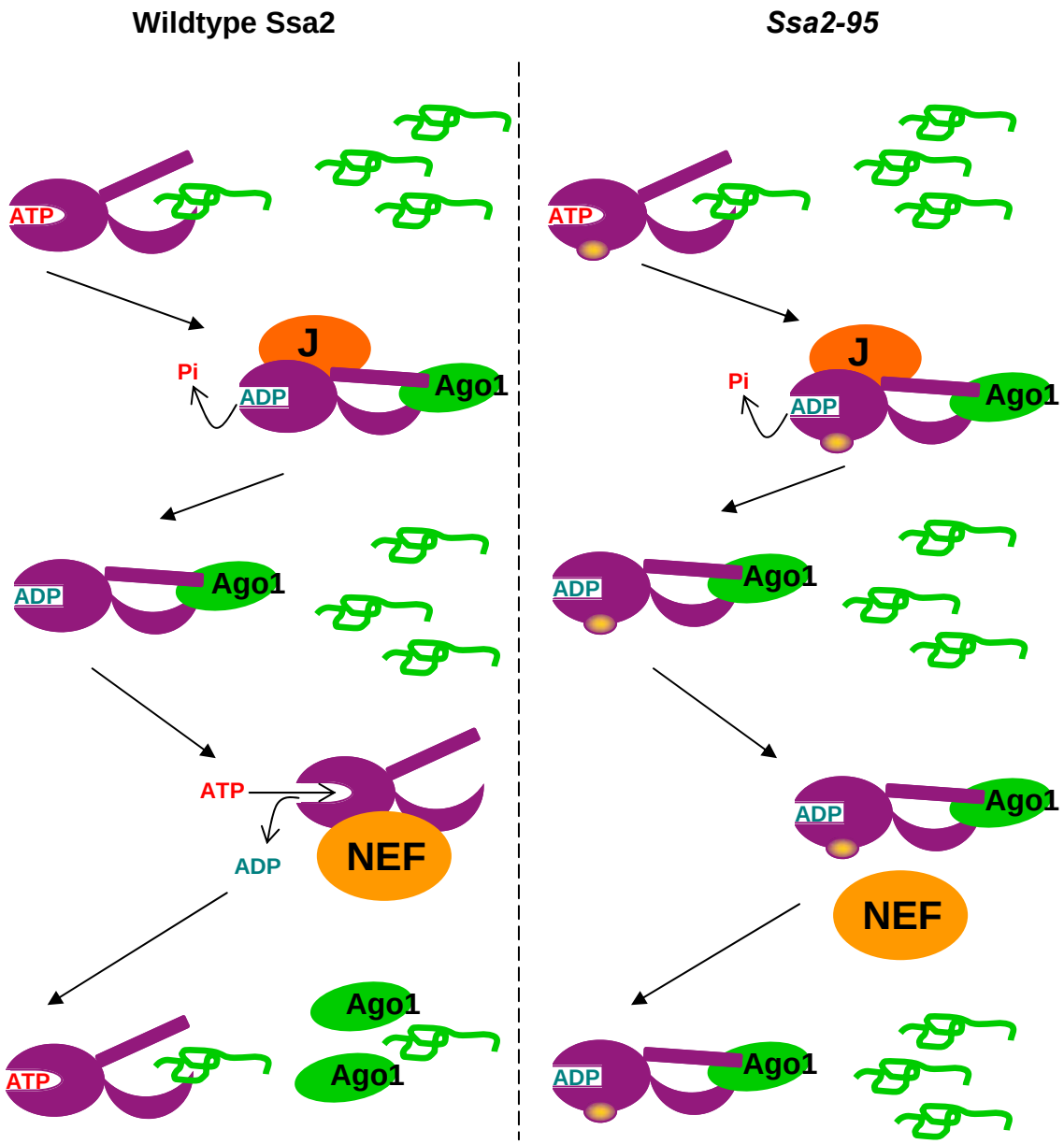


Figure 7.2 – Model of the possible affect of the *ssa2-95* mutation on the RNAi pathway. The *ssa2-95* mutation may cause loss of the interaction between the Ssa2 protein and a nucleotide exchange factor (NEF) resulting in both Ssa2 and the RNAi substrate protein becoming sequestered.

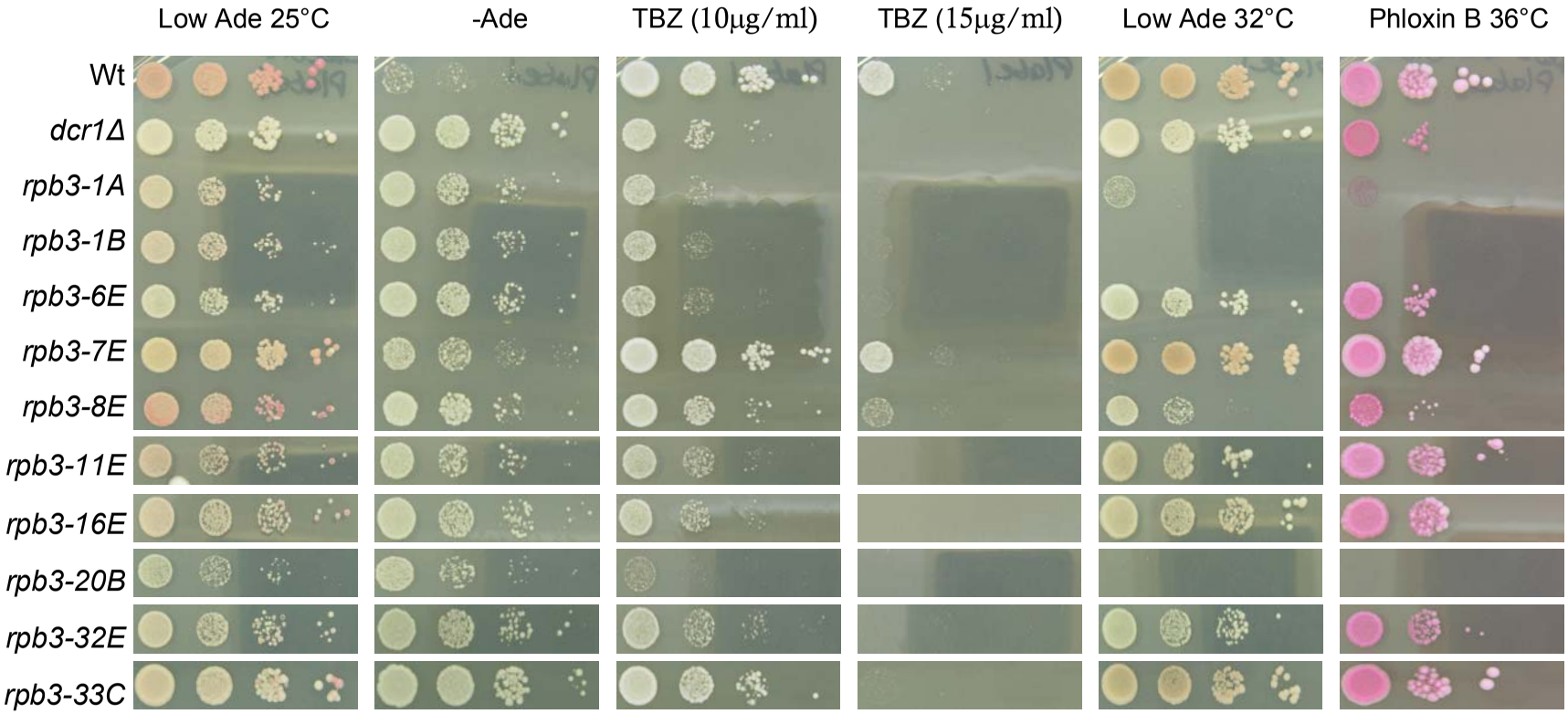
Substrate binding and J proteins stimulate the ATPase activity of the Nucleotide Binding Domain (NBD) of Ssa2 causing hydrolysis of ATP to ADP and a conformational shift close the substrate binding domain (SBD). Nucleotide exchange factors (NEFs) stimulate release of ADP from the NBD allowing consequent binding of ATP and a conformational change to open the SBD.

7.4 Conclusions and perspectives

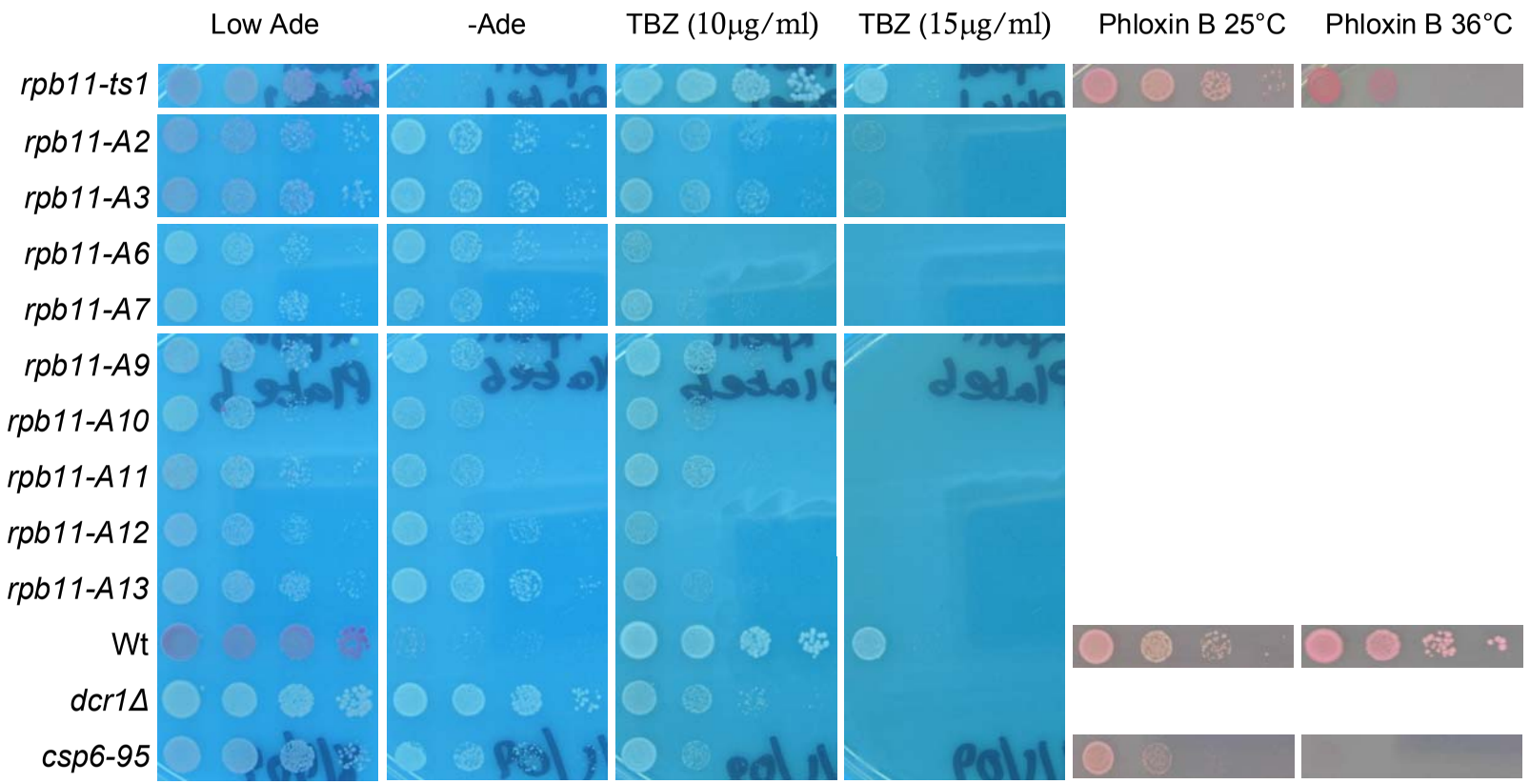
The RNAPII mutants described here have provided some insight into the level of complexity involved in the role(s) of RNAPII in RNAi-directed heterochromatin formation. Analysis of the RNAPII mutants shows that mutations in RNAPII subunits can cause a degree of RNAi-independent silencing of reporter genes. It is tempting to speculate that the RNAi pathway normally recruits heterochromatin forming proteins such as Clr4 by causing a change in transcription which has been inefficiently mimicked by the RNAPII mutations isolated in this study.

I have also shown that *csp6*$^{+}$ is allelic to *ssa2*$^{+}$. The phenotypes of the *ssa2* point mutants suggest that Ssa2 has a role in RNAi-dependent heterochromatin formation, however the fact that *ssa2*Δ has no silencing defects indicates that the different Hsp70 proteins may act redundantly in this pathway. Although heat-shock proteins have been implicated in Ago1 loading in other species, the *ssa2* mutants do not appear to affect the loading efficiency of siRNA into Ago1/RITS in *S. pombe*. As the *ssa2* mutants have a dominant affect on RNAi-dependent heterochromatin formation it is likely that the mutated Hsp70 is capturing a protein/complex required for RNAi and not efficiently releasing it.

Good progress has been made with characterising the roles of a number of proteins with specific roles in the RNAi and chromatin modification pathways. However a full understanding of the mechanism and regulation of RNAi-directed chromatin modification requires that we also understand the interplay between these pathways and other core cellular processes. In this work I have shed light on the complex links between the RNAi machinery and two essential cellular components, RNAPII and heat-shock proteins. Evidence from other species suggests that such links are conserved features of RNAi pathways.



Appendix Figure 1 - Silencing defects and temperature sensitivity of the *rpb3* mutants. Serial dilution comparative plating analyses were used to test the phenotypes. The *ade6*$^{+}$ gene, which is situated in the pericentromeric heterochromatin region, allows silencing to be tested using growth on low adenine (wt is red, defective silencing is white) and minus adenine (no growth for wildtype, good growth for defective silencing) plates. Cells with defective pericentromeric heterochromatin have defects in cohesion between sister chromatids resulting in sensitivity to the microtubule destabilising drug TBZ. To test temperature sensitivity (ts) cells are grown at permissive (25°C) and restrictive (36°C) temperatures on plates supplemented with Phloxin B which stains dead cells red.



Appendix Figure 2 - Silencing defects and temperature sensitivity of the *rpb11* mutants. Serial dilution comparative plating analyses were used to test the phenotypes. The *ade6*$^{+}$ gene, which is situated in the pericentromeric heterochromatin region, allows silencing to be tested using growth on low adenine (wt is red, defective silencing is white) and minus adenine (no growth for wildtype, good growth for defective silencing) plates. Cells with defective pericentromeric heterochromatin have defects in cohesion between sister chromatids resulting in sensitivity to the microtubule destabilising drug TBZ. To test temperature sensitivity (ts) cells are grown at permissive (25°C) and restrictive (36°C) temperatures on plates supplemented with Phloxin B which stains dead cells red.

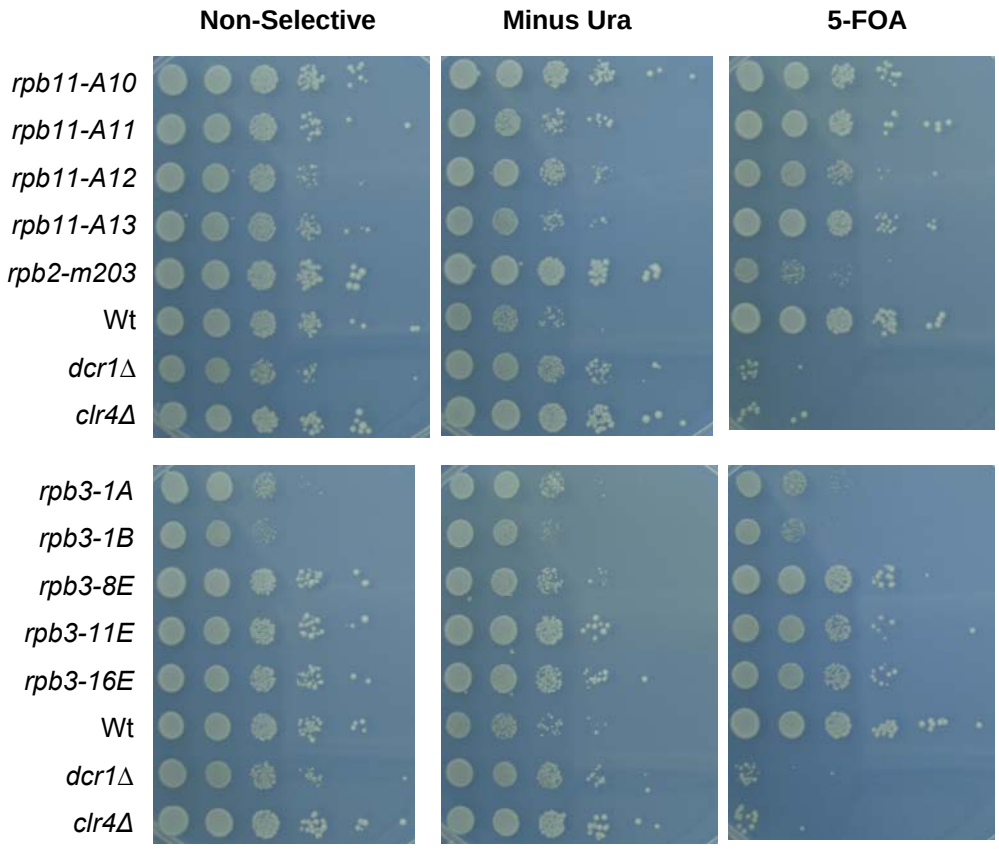


Figure 3 – Serial dilution comparative plating assay testing silencing of the *ura4*$^{+}$ marker gene at the pericentromeric outer repeats of centromere 1. Expression of *ura4*+ causes growth on –URA and death on 5-FOA.

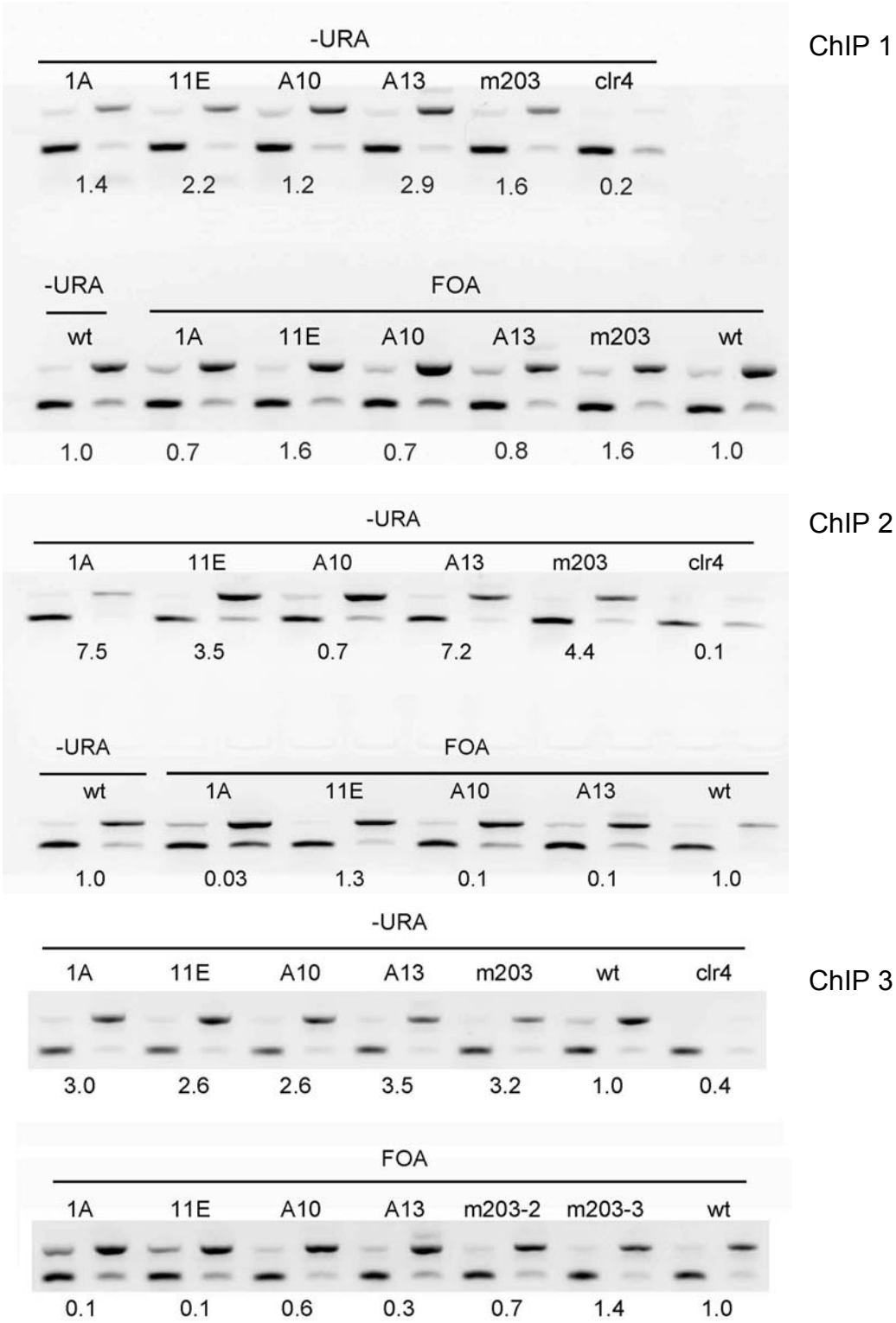


Figure 4 - H3K9me2 levels on dg repeats in the RNAPII mutants when grown selectively in either –URA or 5-FOA to assess variegation. The experiment was carried out in triplicate and the IP'ed DNA was measured using semi-quantitative PCR. The intensity of bands was used to calculate enrichment which was normalised using the fbp gene and the input samples. The enrichments were further normalised for each growth media setting the wt sample as 1.

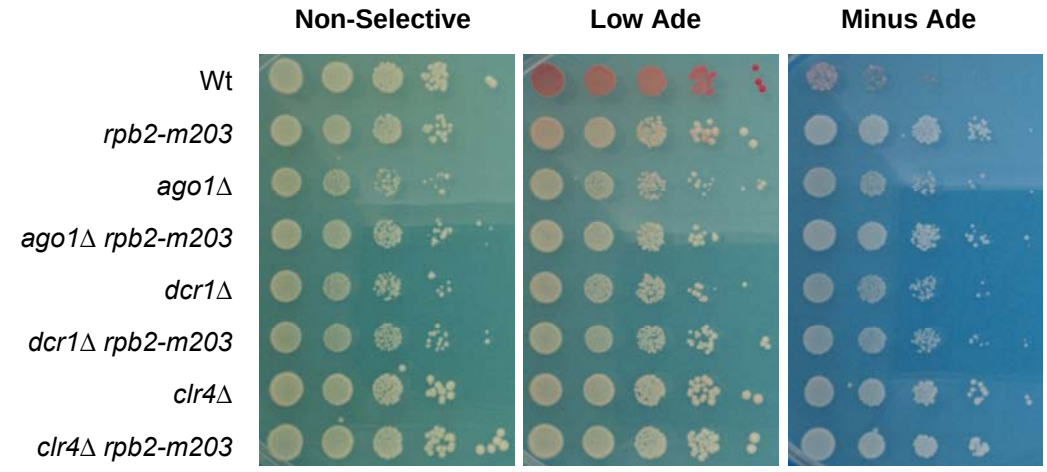


Figure 5 – Serial dilution comparative plating assays testing the silencing of the *ade6*$^{+}$ reporter gene at the pericentromere in *rpb2-m203*-RNAi double mutants. Normally the RNAi mutants *ago1*Δ and *dcr1*Δ lose silencing of reporter genes at the pericentromere and the ClrC mutant *clr4*Δ loses silencing at all heterochromatin loci. *rpb2-m203* does not partially rescue the loss of silencing in *dcr1*Δ and *ago1*Δ unlike the RNAPII mutants *rpb11-A10* and *rpb3-11E*.

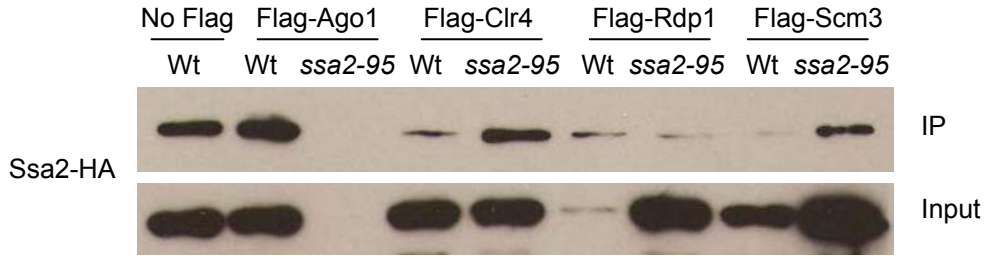


Figure 6 – Testing the interaction of the Ssa2 chaperone with components of the RNAi-mediated heterochromatin formation pathway and the CENP-A chaperone protein Scm3. Initially the Flag-tagged proteins were immunoprecipitated with α-Flag antibody. The pulldowns were then separated by size by SDS-PAGE and the western blot was probed using α-HA antibody to detect the HA-tagged Ssa2 protein.

References cited

Acker, J.l., de Graaff, M., Cheynel, I., Khazak, V., Kedinger, C., and Vigneron, M. (1997). Interactions between the Human RNA Polymerase II Subunits. Journal of Biological Chemistry *272*, 16815-16821.

Aird, D., Ross, M.G., Chen, W.-S., Danielsson, M., Fennell, T., Russ, C., Jaffe, D.B., Nusbaum, C., and Gnirke, A. (2011). Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. Genome Biology *12*, R18-R18.

Allen, T.A., Von Kaenel, S., Goodrich, J.A., and Kugel, J.F. (2004). The SINE-encoded mouse B2 RNA represses mRNA transcription in response to heat shock. Nat Struct Mol Biol *11*, 816-821.

Allshire, R.C., Javerzat, J.-P., Redhead, N.J., and Cranston, G. (1994). Position effect variegation at fission yeast centromeres. Cell *76*, 157-169.

Allshire, R.C., and Karpen, G.H. (2008). Epigenetic regulation of centromeric chromatin: old dogs, new tricks? Nat Rev Genet *9*, 923-937.

Almeida, R., and Allshire, R.C. (2005). RNA silencing and genome regulation. Trends in Cell Biology *15*, 251-258.

Anderson, H.E., Kagansky, A., Wardle, J., Rappsilber, J., Allshire, R.C., and Whitehall, S.K. (2010). Silencing Mediated by the Schizosaccharomyces pombe HIRA Complex Is Dependent upon the Hpc2-Like Protein, Hip4. PLoS ONE *5*, e13488-e13488.

Aravin, A.A., Sachidanandam, R., Girard, A., Fejes-Toth, K., and Hannon, G.J. (2007). Developmentally Regulated piRNA Clusters Implicate MILI in Transposon Control. Science *316*, 744-747.

Armache, K.-J. (2005). Structures of Complete RNA Polymerase II and Its Subcomplex, Rpb4/7. Journal of Biological Chemistry *280*, 7131-7134.

Armache, K.-J., Kettenberger, H., and Cramer, P. (2003). Architecture of initiation-competent 12-subunit RNA polymerase II. Proceedings of the National Academy of Sciences *100*, 6964-6968.

Arney, K.L., and Fisher, A.G. (2004). Epigenetic aspects of differentiation. J Cell Sci *117*, 4355-4363.

Basi, G., Schmid, E., and Maundrell, K. (1993). TATA box mutations in the Schizosaccharomyces pombe nmt1 promoter affect transcription efficiency but not the transcription start point or thiamine repressibility. Gene *123*, 131-136.

Baulcombe, D. (2004). RNA silencing in plants. Nature *431*.

Bayne, E.H., Portoso, M., Kagansky, A., Kos-Braun, I.C., Urano, T., Ekwall, K., Alves, F., Rappsilber, J., and Allshire, R.C. (2008). Splicing Factors Facilitate RNAi-Directed Silencing in Fission Yeast. Science *322*, 602-606.

Bayne, E.H., White, S.A., Kagansky, A., Bijos, D.A., Sanchez-Pulido, L., Hoe, K.-L., Kim, D.-U., Park, H.-O., Ponting, C.P., and Rappsilber, J. (2010). Stc1: A Critical Link between RNAi and Chromatin Modification Required for Heterochromatin Integrity. Cell *140*, 666-677.

Benga, W.J., Grandemange, S., Shpakovski, G.V., Shematorova, E.K., Kedinger, C., and Vigneron, M. (2005). Distinct regions of RPB11 are required for heterodimerization with RPB3 in human and yeast RNA polymerase II. Nucleic Acids Research *33*, 3582-3590.

Bernard, P., Maure, J. F., Partridge, J. F., Genier, S., Javerzat, J. P., Allshire, R. C. (2001). Requirement of Heterochromatin for Cohesion at Centromeres. Science *294*, 2539-2342.

Bijos, D. (2007). Identification and Characterisation of Novel Genes Alleviating Silencing at Centromeres in Fission Yeast. In School of Informatics (University of Edinburgh).

Bjerling, P., Silverstein, R.A., Thon, G., Caudy, A., Grewal, S., and Ekwall, K. (2002). Functional Divergence between Histone Deacetylases in Fission Yeast by Distinct Cellular Localization and In Vivo Specificity. Mol Cell Biol *22*, 2170-2181.

Black, B.E., and Cleveland, D.W. (2011). Epigenetic Centromere Propagation and the Nature of CENP-A Nucleosomes. Cell *144*, 471-479.

Black, B.E., Foltz, D.R., Chakravarthy, S., Luger, K., Woods, V.L., and Cleveland, D.W. (2004). Structural determinants for generating centromeric chromatin. Nature *430*, 578-582.

Blackwell, C., Martin, K.A., Greenall, A., Pidoux, A., Allshire, R.C., and Whitehall, S.K. (2004). The Schizosaccharomyces pombe HIRA-Like Protein Hip1 Is Required for the Periodic Expression of Histone Genes and Contributes to the Function of Complex Centromeres. Mol Cell Biol *24*, 4309-4320.

Blower, M.D., Sullivan, B.A., Karpen, G.H., Laboratory, C.B., Jolla, L., and Diego, S. (2002). Conserved Organization of Centromeric Chromatin in Flies and Humans. Developmental Cell *2*, 319-330.

Blumenstiel, J.P., Noll, A.C., Griffiths, J.A., Perera, A.G., Walton, K.N., Gilliland, W.D., Hawley, R.S., and Staehling-Hampton, K. (2009). Identification of EMS-Induced Mutations in Drosophila melanogaster by Whole-Genome Sequencing. Genetics *182*, 25-32.

Boeke, J.D., LaCroute, F., and Fink, G.R. (1984). A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. Mol Gen Genet *197*, 345-346.

Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G.J. (2007). Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in Drosophila. Cell *128*, 1089-1103.

Brown, S.W. (1966). Heterochromatin. Science *151*, 417-425.

Buhler, M., Haas, W., Gygi, S.P., and Moazed, D. (2007). RNAi-dependent and -independent RNA turnover mechanisms contribute to heterochromatic gene silencing. Cell *129*, 707-721.

Buhler, M., Verdel, A., and Moazed, D. (2006). Tethering RITS to a Nascent Transcript Initiates RNAi- and Heterochromatin-Dependent Gene Silencing. Cell *125*, 873-886.

Buker, S.M., Iida, T., Buhler, M., Villen, J., Gygi, S.P., Nakayama, J.-I., and Moazed, D. (2007). Two different Argonaute complexes are required for siRNA generation and heterochromatin assembly in fission yeast. Nat Struct Mol Biol *14*, 200-207.

Buratowski, S. (2009). Progression through the RNA polymerase II CTD cycle. Molecular Cell *36*, 541-546.

Cam, H.P., Sugiyama, T., Chen, E.S., Chen, X., FitzGerald, P.C., and Grewal, S.I.S. (2005). Comprehensive analysis of heterochromatin- and RNAi-mediated epigenetic control of the fission yeast genome. Nat Genet *37*, 809-819.

Camahort, R., Shivaraju, M., Mattingly, M., Li, B., Nakanishi, S., Zhu, D., Shilatifard, A., Workman, J.L., and Gerton, J.L. (2009). Cse4 Is Part of an Octameric Nucleosome in Budding Yeast. Molecular Cell *35*, 794-805.

Cassidy, S.B., Lai, L.W., Erickson, R.P., Magnuson, L., Thomas, E., Gendron, R., and Herrmann, J. (1992). Trisomy 15 with loss of the paternal 15 as a cause of Prader-Willi syndrome due to maternal disomy. American Journal of Human Genetics *51*, 701-708.

Cheeseman, I.M., and Desai, A. (2008). Molecular architecture of the kinetochore-microtubule interface. Nat Rev Mol Cell Biol *9*, 33-46.

Chen, E.S., Zhang, K., Nicolas, E., Cam, H.P., Zofall, M., and Grewal, S.I.S. (2008). Cell cycle control of centromeric repeat transcription and heterochromatin assembly. Nature *451*, 734-737.

Choi, E.S., Stralfors, A., Castillo, A.G., Durand-Dubief, M., Ekwall, K., and Allshire, R.C. (2011). Identification of non-coding transcripts from within

CENP-A chromatin at fission yeast centromeres. Journal of Biological Chemistry *286*, 23600-23607.

Collins, K.A., Castillo, A.R., Tatsutani, S.Y., and Biggins, S. (2005). De Novo Kinetochore Assembly Requires the Centromeric Histone H3 Variant. Mol Biol Cell *16*, 5649-5660.

Curcio, M.J., and Derbyshire, K.M. (2003). The outs and ins of transposition: from Mu to Kangaroo. Nat Rev Mol Cell Biol *4*, 865-877.

Cyr, D.M. (2008). Swapping Nucleotides, Tuning Hsp70. Cell *133*, 945-947.

Dalal, Y., Wang, H., Lindsay, S., and Henikoff, S. (2007). Tetrameric Structure of Centromeric Nucleosomes in Interphase Drosophila Cells. PLoS Biol *5*, e218-e218.

Daniel, S., Söti, C., Csermely, P., Bradley, G., and Blatch, G.L. (2007). Hop: An Hsp70/Hsp90 Co-Chaperone That Functions Within and Beyond Hsp70/Hsp90 Protein Folding Pathways. In Networking of Chaperones by Co-Chaperones, pp. 26-37.

De Angelis, R., Iezzi, S., Bruno, T., Corbi, N., Di Padova, M., Floridi, A., Fanciulli, M., and Passananti, C. (2003). Functional interaction of the subunit 3 of RNA polymerase II (RPB3) with transcription factor-4 (ATF4). FEBS Letters *547*, 15-19.

de la Mata, M., Alonso, C.R., Kadener, S.n., Fededa, J.P., Blaustein, M., Pelisch, F., Cramer, P., Bentley, D., and Kornblihtt, A.R. (2003). A Slow RNA Polymerase II Affects Alternative Splicing In Vivo. Molecular Cell *12*, 525-532.

DeBeauchamp, J.L., Moses, A., Noffsinger, V.J.P., Ulrich, D.L., Job, G., Kosinski, A.M., and Partridge, J.F. (2008). Chp1-Tas3 Interaction Is Required To Recruit RITS to Fission Yeast Centromeres and for Maintenance of Centromeric Heterochromatin. Mol Cell Biol *28*, 2154-2166.

Djupedal, I., and Ekwall, K. (2008). The Paradox of Silent Heterochromatin. Science, 624-625.

Djupedal, I., Kos-Braun, I.C., Mosher, R.A., Soderholm, N., Simmer, F., Hardcastle, T.J., Fender, A., Heidrich, N., Kagansky, A., Bayne, E.*, et al.* (2009). Analysis of small RNA in fission yeast; centromeric siRNAs are potentially generated through a structured RNA. EMBO J *28*, 3832-3844.

Djupedal, I., Portoso, M., Spåhr, H., Bonilla, C., Gustafsson, C. M., Allshire, R. C., Ekwall, K. (2005). RNA Pol II subunit Rpb7 promotes centromeric transcription and RNAi-directed chromatin silencing. Genes and Development *19*, 2301-2306.

Donaldson, I.M., and Friesen, J.D. (2000). Zinc Stoichiometry of Yeast RNA Polymerase II and Characterization of Mutations in the Zinc-binding Domain of the Largest Subunit. Journal of Biological Chemistry *275*, 13780-13788.

Donze, D., and Kamakaka, R.T. (2001). RNA polymerase III and RNA polymerase II promoter complexes are heterochromatin barriers in Saccharomyces cerevisiae. EMBO J *20*, 520-531.

Dougherty, W.G., and Parks, T.D. (1995). Transgene and gene suppression: telling us something new? Current Opinion in Cell Biology *7*, 399-405.

Dunleavy, E.M., Roche, D.l., Tagami, H., Lacoste, N., Ray-Gallet, D., Nakamura, Y., Daigo, Y., Nakatani, Y., and Almouzni-Pettinotti, G.v. (2009). HJURP Is a Cell-Cycle-Dependent Maintenance and Deposition Factor of CENP-A at Centromeres. Cell *137*, 485-497.

Egel, R. (1984). Two tightly linked silent cassettes in the mating-type region of Schizosaccharomyces pombe. Current Genetics *8*, 199-203.

Egel, R. (2003). The Molecular Biology of Schizosaccharocymes pombe (Springer-Verlag).

Eissenberg, J.C. (1989). Position effect variegation in Drosophila: Towards a genetics of chromatin assembly. BioEssays *11*, 14-17.

Ekwall, K., Cranston, G., Allshire, R. C. (1999). Fission Yeast Mutants That Alleviate Transcriptional Silencing in Centromeric Flanking Repeats and Disrupt Chromosome Segregation. Genetics Society of America *153*, 1153-1169.

Ekwall, K., and Ruusala, T. (1994). Mutations in rik1, clr2, clr3 and clr4 Genes Asymmetrically Derepress the Silent Mating-Type Loci in Fission Yeast. Genetics *136*, 53-64.

Elbashir, S.M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., and Tuschl, T. (2001). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature *411*, 494-498.

Ellermeier, C., Higuchi, E.C., Phadnis, N., Holm, L., Geelhood, J.L., Thon, G., and Smith, G.R. (2010). RNAi and heterochromatin repress centromeric meiotic recombination. Proceedings of the National Academy of Sciences *107*, 8701-8705.

Espinoza, C.A., Allen, T.A., Hieb, A.R., Kugel, J.F., and Goodrich, J.A. (2004). B2 RNA binds directly to RNA polymerase II to repress transcript synthesis. Nat Struct Mol Biol *11*, 822-829.

Eun, S.H., Gan, Q., and Chen, X. (2010). Epigenetic regulation of germ cell differentiation. Current Opinion in Cell Biology *22*, 737-743.

Fabian, M.R., Sonenberg, N., and Filipowicz, W. (2010). Regulation of mRNA Translation and Stability by microRNAs. Annual Review of Biochemistry *79*, 351-379.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature *391*, 806-811.

Fischer, T.s., Cui, B., Dhakshnamoorthy, J., Zhou, M., Rubin, C., Zofall, M., Veenstra, T.D., and Grewal, S.I.S. (2009). Diverse roles of HP1 proteins in heterochromatin assembly and functions in fission yeast. Proceedings of the National Academy of Sciences.

Fischle, W., Wang, Y., and Allis, C., D (2003). Histone and chromatin cross-talk. Current Opinion in Cell Biology *15*, 172-183.

Fisher, C.R. (1969). Enzymology of the pigmented adenine-requiring mutants of Saccharomyces and Schizosaccharomyces. Biochemical and Biophysical Research Communications *34*, 306-310.

Folco, H.D., Pidoux, A.L., Urano, T., and Allshire, R.C. (2008). Heterochromatin and RNAi Are Required to Establish CENP-A Chromatin at Centromeres. Science *319*, 94-97.

Foltz, D.R., Jansen, L.E.T., Bailey, A.O., Yates, J.R., Bassett, E.A., Wood, S., Black, B.E., and Cleveland, D.W. (2009). Centromere-Specific Assembly of CENP-A Nucleosomes Is Mediated by HJURP. Cell *137*, 472-484.

Fong, Y.W., and Zhou, Q. (2001). Stimulatory effect of splicing factors on transcriptional elongation. Nature *414*, 929-933.

Freeman-Cook, L.L., Gómez, E.B., Spedale, E.J., Marlett, J., Forsburg, S.L., Pillus, L., and Laurenson, P. (2005). Conserved Locus-Specific Silencing Functions of Schizosaccharomyces pombe sir2+. Genetics *169*, 1243-1260.

Gavin, A.-C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.-M., Cruciat, C.-M.*, et al.* (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature *415*, 141-147.

Gourse, R.L., Ross, W., and Gaal, T. (2000). UPs and downs in bacterial transcription initiation: the role of the alpha subunit of RNA polymerase in promoter recognition. Molecular Microbiology *37*, 687-695.

Greenall, A., Williams, E.S., Martin, K.A., Palmer, J.M., Gray, J., Liu, C., and Whitehall, S.K. (2006). Hip3 Interacts with the HIRA Proteins Hip1 and Slm9 and Is Required for Transcriptional Silencing and Accurate Chromosome Segregation. Journal of Biological Chemistry *281*, 8732-8739.

Grewal, S.I.S., Bonaduce, M.J., and Klar, A.J.S. (1998). Histone Deacetylase Homologs Regulate Epigenetic Inheritance of Transcriptional Silencing and Chromosome Segregation in Fission Yeast. Genetics *150*, 563-576.

Guang, S., Bochner, A.F., Burkhart, K.B., Burton, N., Pavelec, D.M., and Kennedy, S. (2010). Small regulatory RNAs inhibit RNA polymerase II during the elongation phase of transcription. Nature *465*, 1097-1101.

Gunawardane, L.S., Saito, K., Nishida, K.M., Miyoshi, K., Kawamura, Y., Nagami, T., Siomi, H., and Siomi, M.C. (2007). A Slicer-Mediated Mechanism for Repeat-Associated siRNA 5' End Formation in Drosophila. Science *315*, 1587-1590.

Habara, Y., Urushiyama, S., Shibuya, T., Ohshima, Y., and Tani, T. (2001). Mutation in the prp12+ gene encoding a homolog of SAP130/SF3b130 causes differential inhibition of pre-mRNA splicing and arrest of cell-cycle progression in Schizosaccharomyces pombe. RNA *7*, 671-681.

Hahn, S. (2004). Structure and mechanism of the RNA Polymerase II transcription machinery. Nature structural & molecular biology *11*, 394-403.

Halic, M., and Moazed, D. (2010). Dicer-Independent Primal RNAs Trigger RNAi and Heterochromatin Formation. Cell *140*, 504-516.

Hall, I.M., Shankaranarayana, G.D., Noma, K.-I., Ayoub, N., Cohen, A., and Grewal, S.I.S. (2002). Establishment and maintenance of a heterochromatin domain. Science (New York, NY) *297*, 2232-2237.

Hartl, F.U., and Hayer-Hartl, M. (2002). Molecular Chaperones in the Cytosol: from Nascent Chain to Folded Protein. Science *295*, 1852-1858.

Henikoff, S., and Dalal, Y. (2005). Centromeric chromatin: what makes it unique? Current Opinion in Genetics & Development *15*, 177-184.

Herr, A.J., Jensen, M.B., Dalmay, T., and Baulcombe, D.C. (2005). RNA polymerase IV directs silencing of endogenous DNA. Science *308*, 118-120.

Horn, P.J., Bastie, J.-N., and Peterson, C.L. (2005). A Rik1-associated, cullin-dependent E3 ubiquitin ligase is essential for heterochromatin formation. Genes & Development *19*, 1705-1714.

Iida, T., Kawaguchi, R., and Nakayama, J.-i. (2006). Conserved Ribonuclease, Eri1, Negatively Regulates Heterochromatin Assembly in Fission Yeast. Current Biology *16*, 1459-1464.

Iki, T., Yoshikawa, M., Nishikiori, M., Jaudal, M.C., Matsumoto-Yokoyama, E., Mitsuhara, I., Meshi, T., and Ishikawa, M. (2010). In Vitro Assembly of Plant RNA-Induced Silencing Complexes Facilitated by Molecular Chaperone HSP90. Molecular Cell.

Irvine, D.V., Goto, D.B., Vaughn, M.W., Nakaseko, Y., McCombie, W.R., Yanagida, M., and Martienssen, R. (2009). Mapping epigenetic mutations in fission yeast using whole-genome next-generation sequencing. Genome Research *19*, 1077-1083.

Irvine, D.V., Zaratiegui, M., Tolia, N.H., Goto, D.B., Chitwood, D.H., Vaughn, M.W., Joshua-Tor, L., and Martienssen, R.A. (2006). Argonaute Slicing Is Required for Heterochromatic Silencing and Spreading. Science *313*, 1134-1137.

Ishihama, A. (1981). Subunit of assembly of Escherichia coli RNA polymerase. Advances in Biophysics *14*, 1-35.

Iwasaki, S., Kobayashi, M., Yoda, M., Sakaguchi, Y., Katsuma, S., Suzuki, T., and Tomari, Y. (2010). Hsc70/Hsp90 Chaperone Machinery Mediates ATP-Dependent RISC Loading of Small RNA Duplexes. Molecular Cell.

Javerzat, J.-P., Cranston, G., and Allshire, R.C. (1996). Fission Yeast Genes which Disrupt Mitotic Chromosome Segregation When Overexpressed. Nucleic Acids Research *24*, 4676-4683.

Jenuwein, T., and Allis, C.D. (2001). Translating the Histone Code. Science *293*, 1074-1080.

Jia, S., Noma, K., and Grewal, S. (2004). RNAi-Independent Heterochromatin Nucleation by the Stress-Activated ATF/CREB Family Proteins. Science *304*, 1971-1976.

Jiang, J., Lafer, E.M., and Sousa, R. (2006). Crystallization of a functionally intact Hsc70 chaperone. Acta Crystallographica Section F: Structural Biology and Crystallization Communications *62*, 39–43-39–43.

Jiang, J., Maes, E.G., Taylor, A.B., Wang, L., Hinck, A.P., Lafer, E.M., and Sousa, R. (2007). Structural Basis of J Cochaperone Binding and Regulation of Hsp70. Molecular Cell *28*, 422-433.

Johnston, M., Geoffroy, M.-C., Sobala, A., Hay, R., and Hutvagner, G. (2010). HSP90 Protein Stabilizes Unloaded Argonaute Complexes and Microscopic P-bodies in Human Cells. Mol Biol Cell *21*, 1462-1469.

Jones, L., Ratcliff, F., and Baulcombe, D.C. (2001). RNA-directed transcriptional gene silencing in plants can be inherited independently of the RNA trigger and requires Met1 for maintenance. Current Biology *11*, 747-757.

Kagansky, A., Folco, H.D., Almeida, R., Pidoux, A.L., Boukaba, A., Simmer, F., Urano, T., Hamilton, G.L., and Allshire, R.C. (2009). Synthetic Heterochromatin Bypasses RNAi and Centromeric Repeats to Establish Functional Centromeres. Science *324*, 1716-1719.

Kalantry, S. (2011). Recent advances in X-chromosome inactivation. Journal of Cellular Physiology *226*, 1714-1718.

Kampinga, H.H., and Craig, E.A. (2010). The HSP70 chaperone machinery: J proteins as drivers of functional specificity. Nat Rev Mol Cell Biol *11*, 579-592.

Kanoh, J., Sadaie, M., Urano, T., and Ishikawa, F. (2005). Telomere Binding Protein Taz1 Establishes Swi6 Heterochromatin Independently of RNAi at Telomeres. Current Biology *15*, 1808-1819.

Karpen, G.H., and Allshire, R.C. (1997). The case for epigenetic effects on centromere identity and function. Trends in Genetics *13*, 489-496.

Kato, H., Goto, D. B., Martienssen, R. A., Urano, T., Furukawa, K., Murakami, Y. (2005). RNA Polymerase II is Required for RNAi-Dependent Heterochromatin Assembly. Science *309*, 467-469.

Kavi, H.H., and Birchler, J.A. (2009). Interaction of RNA polymerase II and the small RNA machinery affects heterochromatic silencing in Drosophila. Epigenetics Chromatin *2*, 15-15.

Kennerdell, J.R., and Carthew, R.W. (1998). Use of dsRNA-Mediated Genetic Interference to Demonstrate that frizzled and frizzled 2 Act in the Wingless Pathway. Cell *95*, 1017-1026.

Kettenberger, H., Armache, K.-J., and Cramer, P. (2003). Architecture of the RNA Polymerase II-TFIIS Complex and Implications for mRNA Cleavage. Cell *114*, 347-357.

Kettenberger, H., Armache, K.-J., and Cramer, P. (2004). Complete RNA Polymerase II Elongation Complex Structure and Its Interactions with NTP and TFIIS. Molecular Cell *16*, 955-965.

Kidwell, M.G., and Lisch, D. (1997). Transposable elements as sources of variation in animals and plants. Proceedings of the National Academy of Sciences of the United States of America *94*, 7704-7711.

Kim, D.-U., Hayles, J., Kim, D., Wood, V., Park, H.-O., Won, M., Yoo, H.-S., Duhig, T., Nam, M., Palmer, G.*, et al.* (2010). Analysis of a genome-wide set of gene deletions in the fission yeast Schizosaccharomyces pombe. Nat Biotech *28*, 617-623.

Kim, H.S., Choi, E.S., Shin, J.A., Jang, Y.K., and Park, S.D. (2004). Regulation of Swi6/HP1-dependent Heterochromatin Assembly by Cooperation of Components of the Mitogen-activated Protein Kinase Pathway and a Histone Deacetylase Clr6. Journal of Biological Chemistry *279*, 42850-42859.

Kimura, M., Ishiguro, A., and Ishihama, A. (1997). RNA Polymerase II Subunits 2, 3, and 11 Form a Core Subassembly with DNA Binding Activity. Journal of Biological Chemistry *272*, 25851-25855.

Kloc, A., Zaratiegui, M., Nora, E., and Martienssen, R. (2008). RNA Interference Guides Histone Modification during the S Phase of Chromosomal Replication. Current Biology *18*, 490-495.

Klose, R.J., Kallin, E.M., and Zhang, Y. (2006). JmjC-domain-containing proteins and histone demethylation. Nat Rev Genet *7*, 715-727.

Kornberg, R.D. (1974). Chromatin structure: a repeating unit of histones and DNA. Science (New York, NY) *184*, 868-871.

Kugel, J.F., and Goodrich, J.A. (2002). Translocation after synthesis of a four-nucleotide RNA commits RNA polymerase II to promoter escape. Molecular and Cellular Biology *22*, 762-773.

Kuramochi-Miyagawa, S., Watanabe, T., Gotoh, K., Totoki, Y., Toyoda, A., Ikawa, M., Asada, N., Kojima, K., Yamaguchi, Y., Ijiri, T.W.*, et al.* (2008). DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. Genes & Development *22*, 908-917.

Kwek, K.Y., Murphy, S., Furger, A., Thomas, B., O'Gorman, W., Kimura, H., Proudfoot, N.J., and Akoulitchev, A. (2002). U1 snRNA associates with TFIIH and regulates transcriptional initiation. Nature Structural Biology.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W.*, et al.* (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Lejeune, J., Gautier, M., and Turpin, R. (1959). [Study of somatic chromosomes from 9 mongoloid children.]. Comptes Rendus Hebdomadaires Des S,ances De l'Acad,mie Des Sciences *248*, 1721-1722.

Li, B., Carey, M., and Workman, J.L. (2007). The Role of Chromatin during Transcription. Cell *128*, 707-719.

Li, C.F. (2006). An ARGONAUTE4-containing nuclear processing center colocalized with Cajal bodies in Arabidopsis thaliana. Cell *126*, 93-106.

Li, F., Martienssen, R., and Cande, W.Z. (2011). Coordination of DNA replication and histone modification by the Rik1-Dos2 complex. Nature *475*, 244-248.

Liberek, K., Marszalek, J., Ang, D., Georgopoulos, C., and Zylicz, M. (1991). Escherichia coli DnaJ and GrpE heat shock proteins jointly stimulate ATPase activity of DnaK. Proceedings of the National Academy of Sciences *88*, 2874-2878.

Lingel, A., Simon, B., Izaurralde, E., and Sattler, M. (2004). Nucleic acid 3[prime]-end recognition by the Argonaute2 PAZ domain. Nat Struct Mol Biol *11*, 576-577.

Liu, Q., and Hendrickson, W.A. (2007). Insights into Hsp70 Chaperone Activity from a Crystal Structure of the Yeast Hsp110 Sse1. Cell *131*, 106-120.

Lu, J., and Gilbert, D.M. (2007). Proliferation-dependent and cell cycle regulated transcription of mouse pericentric heterochromatin. The Journal of Cell Biology *179*, 411-421.

Luco, R.F., Pan, Q., Tominaga, K., Blencowe, B.J., Pereira-Smith, O.M., and Misteli, T. (2010). Regulation of Alternative Splicing by Histone Modifications. Science *327*, 996-1000.

Luo, S.-W., Zhang, C., Zhang, B., Kim, C.-H., Qiu, Y.-Z., Du, Q.-S., Mei, L., and Xiong, W.-C. (2009). Regulation of heterochromatin remodelling and myogenin expression during muscle differentiation by FAK interaction with MBD2. EMBO J *28*, 2568-2582.

Ma, J.-B., Ye, K., and Patel, D.J. (2004). Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. Nature *429*, 318-322.

MacRae, I.J., Zhou, K., Li, F., Repic, A., Brooks, A.N., Cande, W.Z., Adams, P.D., and Doudna, J.A. (2006). Structural Basis for Double-Stranded RNA Processing by Dicer. Science *311*, 195-198.

Mariner, P.D., Walters, R.D., Espinoza, C.A., Drullinger, L.F., Wagner, S.D., Kugel, J.F., and Goodrich, J.A. (2008). Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock. Molecular Cell *29*, 499-509.

Martens, J.H.A., O'Sullivan, R.J., Braunschweig, U., Opravil, S., Radolf, M., Steinlein, P., and Jenuwein, T. (2005). The profile of repeat-associated histone lysine methylation states in the mouse epigenome. The EMBO Journal *24*, 800-812.

Matranga, C., Tomari, Y., Shin, C., Bartel, D.P., and Zamore, P.D. (2005). Passenger-Strand Cleavage Facilitates Assembly of siRNA into Ago2-Containing RNAi Enzyme Complexes. Cell *123*, 607-620.

Maundrell, K. (1993). Thiamine-repressible expression vectors pREP and pRIP for fission yeast. Gene *123*, 127-130.

May, B.P., Lippman, Z.B., Fang, Y., Spector, D.L., and Martienssen, R.A. (2005). Differential Regulation of Strand-Specific Transcripts from Arabidopsis Centromeric Satellite Repeats. PLoS Genet *1*, e79-e79.

McEwen, B.F., Hsieh, C.-E., Mattheyses, A.L., and Rieder, C.L. (1998). A new look at kinetochore structure in vertebrate somatic cells using high-pressure freezing and freeze substitution. Chromosoma *107*, 366-375.

Meister, G., and Tuschl, T. (2004). Mechanisms of gene silencing by double-stranded RNA. Nature *431*, 343-349.

Mitobe, J., Mitsuzawa, H., Yasui, K., and Ishihama, A. (1999). Isolation and characterization of temperature-sensitive mutations in the gene (rpb3) for subunit 3 of RNA polymerase II in the fission yeast Schizosaccharomyces pombe. Molecular General Genetics *262*, 73-84.

Miyoshi, T., Takeuchi, A., Siomi, H., and Siomi, M.C. (2010). A direct role for Hsp90 in pre-RISC formation in Drosophila. Nat Struct Mol Biol *17*, 1024-1026.

Mizuguchi, G., Xiao, H., Wisniewski, J., Smith, M.M., and Wu, C. (2007). Nonhistone Scm3 and Histones CenH3-H4 Assemble the Core of Centromere-Specific Nucleosomes. Cell *129*, 1153-1164.

Motamedi, M.R., Colmenares, S.U., Gerber, S.A., Gygi, S.P., Moazed, D., and Facility, T.B.M.S. (2004). Two RNAi Complexes, RITS and RDRC, Physically Interact and Localize to Noncoding Centromeric RNAs. Cell *119*, 789-802.

Motamedi, M.R., Hong, E.-J.E., Li, X., Gerber, S., Denison, C., Gygi, S., and Moazed, D. (2008). HP1 Proteins Form Distinct Complexes and Mediate Heterochromatic Gene Silencing by Nonoverlapping Mechanisms. Molecular Cell *32*, 778-790.

Mourelatos, Z., Dostie, J.e., Paushkin, S., Sharma, A., Charroux, B., Abel, L., Rappsilber, J., Mann, M., and Dreyfuss, G. (2002). miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. Genes & Development *16*, 720-728.

Musacchio, A., and Salmon, E.D. (2007). The spindle-assembly checkpoint in space and time. Nat Rev Mol Cell Biol *8*, 379-393.

Nakayama, J.-i., Rice, J.C., Strahl, B.D., Allis, C.D., and Grewal, S.I.S. (2001). Role of Histone H3 Lysine 9 Methylation in Epigenetic Control of Heterochromatin Assembly. Science *292*, 110-113.

Noma, K.-i., Cam, H.P., Maraia, R.J., and Grewal, S.I.S. (2006). A Role for TFIIIC Transcription Factor Complex in Genome Organization. Cell *125*, 859-872.

Noma, K.-i., Sugiyama, T., Cam, H., Verdel, A., Zofall, M., Jia, S., Moazed, D., and Grewal, S.I.S. (2004). RITS acts in cis to promote RNA interference-mediated transcriptional and post-transcriptional silencing. Nat Genet *36*, 1174-1180.

Onodera, Y., Haag, J.R., Ream, T., Nunes, P.C., Pontes, O., and Pikaard, C.S. (2005). Plant Nuclear RNA Polymerase IV Mediates siRNA and DNA Methylation-Dependent Heterochromatin Formation. Cell *120*, 613-622.

Otto, H., Conz, C., Maier, P., WÃ¶lfle, T., Suzuki, C.K., JenÃ¶, P., RÃ¼cknagel, P., Stahl, J., and Rospert, S. (2005). The chaperones MPP11 and Hsp70L1 form the mammalian ribosome-associated complex. Proceedings of the National Academy of Sciences of the United States of America *102*, 10064-10069.

Oudet, P., Gross-Bellard, M., and Chambon, P. (1975). Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. Cell *4*, 281-300.

Palangat, M., Renner, D.B., Price, D.H., and Landick, R. (2005). A negative elongation factor for human RNA polymerase II inhibits the anti-arrest transcript-cleavage factor TFIIS. Proceedings of the National Academy of Sciences of the United States of America *102*, 15036-15041.

Partridge, J., Debeauchamp, J., Kosinski, A., Ulrich, D., Hadler, M., and Noffsinger, V. (2007). Functional Separation of the Requirements for Establishment and Maintenance of Centromeric Heterochromatin. Molecular Cell *26*, 593-602.

Partridge, J.F., Borgstrom, B., and Allshire, R.C. (2000). Distinct protein interaction domains and protein spreading in a complex centromere. Genes Dev *14*, 783-791.

Partridge, J.F., Scott, K.S.C., Bannister, A.J., Kouzarides, T., and Allshire, R.C. (2002). cis-Acting DNA from Fission Yeast Centromeres Mediates Histone H3 Methylation and Recruitment of Silencing Factors and Cohesin to an Ectopic Site. Current Biology *12*, 1652-1660.

Pellman, D. (2007). Cell biology: Aneuploidy and cancer. Nature *446*, 38-39.

Picard, D. (2011). List of Hsp90 Interactors, www.picard.ch/downloads/Hsp90interactors.pdf (Accessed 29/04/2011).

Pidoux, A.L., Choi, E.S., Abbott, J.K.R., Liu, X., Kagansky, A., Castillo, A.G., Hamilton, G.L., Richardson, W., Rappsilber, J., and He, X. (2009). Fission Yeast Scm3: A CENP-A Receptor Required for Integrity of Subkinetochore Chromatin. Molecular Cell *33*, 299-311.

Polizzi, C., and Clarke, L. (1991). The chromatin structure of centromeres from fission yeast: differentiation of the central core that correlates with function. J Cell Biol *112*, 191-201.

Ponicsan, S.L., Kugel, J.F., and Goodrich, J.A. (2010). Genomic gems: SINE RNAs regulate mRNA production. Current opinion in genetics & development *20*, 149-155.

Portoso, M. (2005). Silent Chromatin Formation and Histone Modifications in Fission Yeast. (University of Edinburgh).

Pratt, W.B., and Toft, D.O. (2003). Regulation of Signaling Protein Function and Trafficking by the hsp90/hsp70-Based Chaperone Machinery. Exp Biol Med *228*, 111-133.

Richter, K., Haslbeck, M., and Buchner, J. (2010). The Heat Shock Response: Life on the Verge of Death. Molecular Cell *40*, 253-266.

Roguev, A., Shevchenko, A., Schaft, D., Thomas, H., Stewart, A.F., and Shevchenko, A. (2004). A Comparative Analysis of an Orthologous Proteomic Environment in the Yeasts Saccharomyces cerevisiae and Schizosaccharomyces pombe. Molecular & Cellular Proteomics *3*, 125-132.

Romano, N., and Macino, G. (1992). Quelling: transient inactivation of gene expression in Neurospora crassa by transformation with homologous sequences. Molecular Microbiology *6*, 3343-3353.

Rudert, F., Bronner, S., Garnier, J.M., and DollÃ©, P. (1995). Transcripts from opposite strands of gamma satellite DNA are differentially expressed during mouse development. Mammalian Genome: Official Journal of the International Mammalian Genome Society *6*, 76-83.

Rudiger, S., Germeroth, L., Schneider-Mergener, J., and Bukau, B. (1997). Substrate specificity of the DnaK chaperone determined by screening cellulose-bound peptide libraries. EMBO J *16*, 1501-1507.

Sadaie, M., Iida, T., Urano, T., and Nakayama, J.-i. (2004). A chromodomain protein, Chp1, is required for the establishment of heterochromatin in fission yeast. EMBO J *23*, 3825-3835.

Saito, K., Nishida, K.M., Mori, T., Kawamura, Y., Miyoshi, K., Nagami, T., Siomi, H., and Siomi, M.C. (2006). Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the Drosophila genome. Genes & Development *20*, 2214-2222.

Sanyal, K., Baum, M., and Carbon, J. (2004). Centromeric DNA sequences in the pathogenic yeast Candida albicans are all different and unique. Proceedings of the National Academy of Sciences of the United States of America *101*, 11374-11379.

Sarin, S., Prabhu, S., O'Meara, M.M., Pe'er, I., and Hobert, O. (2008). Caenorhabditis elegans mutant allele identification by whole-genome sequencing. Nature methods *5*, 865-867.

Schiff, L.A., Nibert, M.L., and Fields, B.N. (1988). Characterization of a zinc blotting technique: evidence that a retroviral gag protein binds zinc. Proceedings of the National Academy of Sciences *85*, 4195-4199.

Scott, K.C., Merrett, S.L., and Willard, H.F. (2006). A Heterochromatin Barrier Partitions the Fission Yeast Centromere into Discrete Chromatin Domains. Current Biology *16*, 119-129.

Seidl, C.I.M., Stricker, S.H., and Barlow, D.P. (2006). The imprinted Air ncRNA is an atypical RNAPII transcript that evades splicing and escapes nuclear export. EMBO J *25*, 3565-3575.

Selth, L.A., Sigurdsson, S., and Svejstrup, J.Q. (2010). Transcript Elongation by RNA Polymerase II. Annual Review of Biochemistry *79*, 271-293.

Shankaranarayana, G.D., Motamedi, M.R., Moazed, D., and Grewal, S.I.S. (2003). Sir2 Regulates Histone H3 Lysine 9 Methylation and Heterochromatin Assembly in Fission Yeast. Current Biology *13*, 1240-1246.

Shen, Y., and Hendershot, L.M. (2005). ERdj3, a Stress-inducible Endoplasmic Reticulum DnaJ Homologue, Serves as a CoFactor for BiP's Interactions with Unfolded Substrates. Mol Biol Cell *16*, 40-50.

Sigova, A., Rhind, N., and Zamore, P.D. (2004). A single Argonaute protein mediates both transcriptional and posttranscriptional silencing in Schizosaccharomyces pombe. Genes & Development *18*, 2359-2367.

Simmer, F., Buscaino, A., Kos-Braun, I.C., Kagansky, A., Boukaba, A., Urano, T., Kerr, A.R.W., and Allshire, R.C. (2010). Hairpin RNA induces secondary small interfering RNA synthesis and silencing in trans in fission yeast. EMBO Rep *11*, 112-118.

Sims, R.J., Millhouse, S., Chen, C.-F., Lewis, B.A., Erdjument-Bromage, H., Tempst, P., Manley, J.L., and Reinberg, D. (2007). Recognition of Trimethylated Histone H3 Lysine 4 Facilitates the Recruitment of Transcription Postinitiation Factors and Pre-mRNA Splicing. Molecular Cell *28*, 665-676.

Siomi, M.C., Sato, K., Pezic, D., and Aravin, A.A. (2011). PIWI-interacting small RNAs: the vanguard of genome defence. Nat Rev Mol Cell Biol *12*, 246-258.

Song, J.-J., Smith, S.K., Hannon, G.J., and Joshua-Tor, L. (2004). Crystal Structure of Argonaute and Its Implications for RISC Slicer Activity. Science *305*, 1434-1437.

Soutourina, J., Wydau, S., Ambroise, Y., Boschiero, C., and Werner, M. (2011). Direct Interaction of RNA Polymerase II and Mediator Required for Transcription in Vivo. Science *331*, 1451-1454.

Sugiyama, T., Cam, H., Sugiyama, R., Noma, K., Zofall, M., Kobayashi, R., and Grewal, S. (2007). SHREC, an Effector Complex for Heterochromatic Transcriptional Silencing. Cell *128*, 491-504.

Sugiyama, T., Cam, H., Verdel, A., Moazed, D., and Grewal, S.I.S. (2005). RNA-dependent RNA polymerase is an essential component of a self-enforcing loop coupling heterochromatin assembly to siRNA production. Proceedings of the National Academy of Sciences of the United States of America *102*, 152-157.

Sullivan, B.A., and Karpen, G.H. (2004). Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin. Nat Struct Mol Biol *11*, 1076-1083.

Szabo, A., Langer, T., Schröder, H., Flanagan, J., Bukau, B., and Hartl, F.U. (1994). The ATP hydrolysis-dependent reaction cycle of the Escherichia coli Hsp70 system DnaK, DnaJ, and GrpE. Proceedings of the National Academy of Sciences of the United States of America *91*, 10345-10349.

Tachiwana, H., Kagawa, W., Shiga, T., Osakabe, A., Miya, Y., Saito, K., Hayashi-Takanaka, Y., Oda, T., Sato, M., Park, S.-Y.*, et al.* (2011). Crystal structure of the human centromeric nucleosome containing CENP-A. Nature *476*, 232-235.

Tahbaz, N., Kolb, F.A., Zhang, H., Jaronczyk, K., Filipowicz, W., and Hobman, T.C. (2004). Characterization of the interactions between mammalian PAZ PIWI domain proteins and Dicer. EMBO reports *5*, 189-194.

Takahashi, K., Murakami, S., Chikashige, Y., Funabiki, H., Niwa, O., and Yanagida, M. (1992). A low copy number central sequence with strict symmetry and unusual chromatin structure in fission yeast centromere. Molecular Biology of the Cell *3*, 819-835.

Tan, Q., Linask, K.L., Ebright, R.H., and Woychik, N.A. (2000). Activation mutants in yeast RNA polymerase II subunit RPB3 provide evidence for a structurally conserved surface required for activation in eukaryotes and bacteria. Genes & Development *14*, 339-348.

Tavaria, M., Gabriele, T., Kola, I., and Anderson, R.L. (1996). A hitchhiker's guide to the human Hsp70 family. Cell Stress & Chaperones *1*, 23-28.

Thon, G., Bjerling, P., Bünner, C.M., and Verhein-Hansen, J. (2002). Expression-state boundaries in the mating-type region of fission yeast. Genetics *161*, 611-622.

Thon, G., and Klar, A.J.S. (1992). The clr1 Locus Regulates the Expression of the Cryptic Mating-Type Loci of Fission Yeast. Genetics *131*, 287-296.

Tomkiewicz, D., Nouwen, N., and Driessen, A.J.M. (2007). Pushing, pulling and trapping - Modes of motor protein supported protein translocation. FEBS Letters *581*, 2820-2828.

Trewick, S.C., Minc, E., Antonelli, R., Urano, T., and Allshire, R.C. (2007). The JmjC domain protein Epe1 prevents unregulated assembly and disassembly of heterochromatin. EMBO J *26*, 4670-4682.

Ulmasov, T., Larkin, R.M., and Guilfoyle, T.J. (1996). Association between 36- and 13.6-kDa -Like Subunits of Arabidopsis thaliana RNA Polymerase II. Journal of Biological Chemistry *271*, 5085-5094.

Vagin, V.V., Sigova, A., Li, C., Seitz, H., Gvozdev, V., and Zamore, P.D. (2006). A Distinct Small RNA Pathway Silences Selfish Genetic Elements in the Germline. Science *313*, 320-324.

Valgardsdottir, R., Chiodi, I., Giordano, M., Cobianchi, F., Riva, S., and Biamonti, G. (2005). Structural and Functional Characterization of Noncoding Repetitive RNAs Transcribed in Stressed Human Cells. Molecular Biology of the Cell *16*, 2597-2604.

Verdel, A. (2004). RNAi-Mediated Targeting of Heterochromatin by the RITS Complex. Science *303*, 672-676.

Vermaak, D., Hayden, H.S., and Henikoff, S. (2002). Centromere Targeting Element within the Histone Fold Domain of Cid. Mol Cell Biol *22*, 7553-7561.

Volpe, T., Schramke, V., Hamilton, G. L., White, S. A., Weng, G., Martienssen, R. A., Allshire, R. C. (2003). RNA interference is required for normal centromere function in fission yeast. Chromosome Research *11*, 137-146.

Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I.S., and Martienssen, R.A. (2002). Regulation of Heterochromatic Silencing and Histone H3 Lysine-9 Methylation by RNAi. Science *297*, 1833-1837.

Warburton, P.E. (2004). Chromosomal dynamics of human neocentromere formation. Chromosome Research *12*, 617-626.

Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P.*, et al.* (2002). Initial sequencing and comparative analysis of the mouse genome. Nature *420*, 520-562.

Welburn, J.P.I., and Cheeseman, I.M. (2008). Toward a Molecular Structure of the Eukaryotic Kinetochore. Developmental Cell *15*, 645-655.

Werner-Washburne, M., Stone, D.E., and Craig, E.A. (1987). Complex interactions among members of an essential subfamily of hsp70 genes in Saccharomyces cerevisiae. Mol Cell Biol *7*, 2568-2577.

Willard, H.F. (1985). Chromosome-specific organization of human alpha satellite DNA. American Journal of Human Genetics *37*, 524-532.

Williams, J.S., Hayashi, T., Yanagida, M., and Russell, P. (2009). Fission Yeast Scm3 Mediates Stable Assembly of Cnp1 (CENP-A) into Centromeric Chromatin. Molecular cell *33*, 287-298.

Wiren, M., Silverstein, R.A., Sinha, I., Walfridsson, J., Lee, H.-m., Laurenson, P., Pillus, L., Robyr, D., Grunstein, M., and Ekwall, K. (2005). Genome wide analysis of nucleosome density histone acetylation and HDAC function in fission yeast. EMBO J *24*, 2906-2918.

Wood, V., Gwilliam, R., Rajandream, M.A., Lyne, M., Lyne, R., Stewart, A., Sgouros, J., Peat, N., Hayles, J., Baker, S.*, et al.* (2002). The genome sequence of Schizosaccharomyces pombe. Nature *415*, 871-880.

Workman, J.L. (2006). Nucleosome displacement in transcription. Genes & Development *20*, 2009-2017.

Yamada, T., Fischle, W., Sugiyama, T., Allis, C., and Grewal, S. (2005). The Nucleation and Maintenance of Heterochromatin by a Histone Deacetylase in Fission Yeast. Molecular Cell *20*, 173-185.

Yamane, K., Mizuguchi, T., Cui, B., Zofall, M., Noma, K.-i., and Grewal, S.I.S. (2011). Asf1/HIRA Facilitate Global Histone Deacetylation and Associate with HP1 to Promote Nucleosome Occupancy at Heterochromatic Loci. Molecular Cell *41*, 56-66.

Young, J.C., Barral, J.M., and Ulrich Hartl, F. (2003). More than folding: localized functions of cytosolic chaperones. Trends in Biochemical Sciences *28*, 541-547.

Zhang, K., Fischer, T., Porter, R.L., Dhakshnamoorthy, J., Zofall, M., Zhou, M., Veenstra, T., and Grewal, S.I.S. (2011). Clr4/Suv39 and RNA Quality Control Factors Cooperate to Trigger RNAi and Suppress Antisense RNA. Science *331*, 1624-1627.

Zhang, K., Mosch, K., Fischle, W., and Grewal, S.I.S. (2008). Roles of the Clr4 methyltransferase complex in nucleation, spreading and maintenance of heterochromatin. Nature Structural & Molecular Biology *15*, 381-388.

Zhu, X., Zhao, X., Burkholder, W.F., Gragerov, A., Ogata, C.M., Gottesman, M.E., and Hendrickson, W.A. (1996). Structural Analysis of Substrate Binding by the Molecular Chaperone DnaK. Science *272*, 1606-1614.

Zofall, M., and Grewal, S.I.S. (2006). Swi6/HP1 Recruits a JmjC Domain Protein to Facilitate Transcription of Heterochromatic Repeats. Molecular Cell *22*, 681-692.

Zofall, M., and Grewal, S.I.S. (2007). HULC, a Histone H2B Ubiquitinating Complex, Modulates Heterochromatin Independent of Histone Methylation in Fission Yeast. Journal of Biological Chemistry *282*, 14065-14072.

Zorio, D.A.R., and Bentley, D.L. (2004). The link between mRNA processing and transcription: communication works both ways. Experimental Cell Research *296*, 91-97.

Acknowledgements

Thank you to everyone in the Allshire Lab for making the Lab such a pleasure to work in. In particular:
Robin go raibh míle maith agat, for all of the sage advice and for sharing your enthusiasm and seemingly crazy yet scarily insightful ideas. A massive thank you to my Lab big sisters Liz and Shaz for putting up with my inane questions and for keeping me on the straight and narrow with all of your excellent advice. Thank you to Ali and Eun for their encyclopaedic knowledge of pombology and to George for being our Lab Mum. Thanks to the tea break crew; Matt, Elisabeth, Anne and Pauline and a particularly big thank you to my good friend Sandra…I hope we will continue to have many more of our enjoyable conversations on and off the topic of science. Thanks to the badminton regulars for keeping me sane through the PhD and especially to Ricardo for our many conversations about science, bikes, sci-fi and of course badminton.

Thanks to the GenePool sequencing service for their genome sequencing expertise. Thanks to Mark Blaxter for allowing me to join his team for training and for providing the *S. japonicus* sequencing free of charge. Thanks to Sujai and Anna for being good hosts and thanks to Urmi for providing her excellent bioinformatics skills. Thank you also to Marian for putting up with the numerous emails asking "are we on the machine yet?"

And last but definitely not least thank you to my family, I have only managed to finish this tome because of the help and support from of each of you:
Rosie you have been a supportive and loving wife, thank you for following me to Edinburgh and going through the trauma, sleepless nights and general stresses with me over the last four years. Thanks to Stew for all of your help, your company and most recently some excellent thesis assembly! Thank you to my loving parents and sister for all of your love, support and encouragement.

Comparative Functional Genomics of the Fission Yeasts

Nicholas Rhind, *et al.*

Science **332**, 930 (2011);
DOI: 10.1126/science.1203357

Comparative Functional Genomics of the Fission Yeasts

Nicholas Rhind,[1]¶ Zehua Chen,[2] Moran Yassour,[3,4,5]‖ Dawn A. Thompson,[3]‖ Brian J. Haas,[2]‖ Naomi Habib,[5,6]‖ Ilan Wapinski,[3,7]‖ Sushmita Roy,[3,8]‖ Michael F. Lin,[8] David I. Heiman,[2] Sarah K. Young,[2] Kanji Furuya,[9] Yabin Guo,[10] Alison Pidoux,[11] Huei Mei Chen,[12] Barbara Robbertse,[13]* Jonathan M. Goldberg,[2] Keita Aoki,[9] Elizabeth H. Bayne,[11]† Aaron M. Berlin,[2] Christopher A. Desjardins,[2] Edward Dobbs,[11] Livio Dukaj,[1] Lin Fan,[2] Michael G. FitzGerald,[2] Courtney French,[6] Sharvari Gujja,[2] Klavs Hansen,[14]‡ Dan Keifenheim,[1] Joshua Z. Levin,[2] Rebecca A. Mosher,[15]§ Carolin A. Müller,[16] Jenna Pfiffner,[2] Margaret Priest,[2] Carsten Russ,[2] Agata Smialowska,[17,18] Peter Swoboda,[17] Sean M. Sykes,[2] Matthew Vaughn,[14] Sonya Vengrova,[19] Ryan Yoder,[13] Qiandong Zeng,[2] Robin Allshire,[11] David Baulcombe,[15] Bruce W. Birren,[20] William Brown,[16] Karl Ekwall,[17,18] Manolis Kellis,[8,3] Janet Leatherwood,[12] Henry Levin,[10] Hanah Margalit,[6] Rob Martienssen,[14] Conrad A. Nieduszynski,[16] Joseph W. Spatafora,[13] Nir Friedman,[5,21] Jacob Z. Dalgaard,[19] Peter Baumann,[22,23,24] Hironori Niki,[9] Aviv Regev,[3,4,24]¶ Chad Nusbaum[2]¶

The fission yeast clade—comprising *Schizosaccharomyces pombe*, *S. octosporus*, *S. cryophilus*, and *S. japonicus*—occupies the basal branch of Ascomycete fungi and is an important model of eukaryote biology. A comparative annotation of these genomes identified a near extinction of transposons and the associated innovation of transposon-free centromeres. Expression analysis established that meiotic genes are subject to antisense transcription during vegetative growth, which suggests a mechanism for their tight regulation. In addition, trans-acting regulators control new genes within the context of expanded functional modules for meiosis and stress response. Differences in gene content and regulation also explain why, unlike the budding yeast of Saccharomycotina, fission yeasts cannot use ethanol as a primary carbon source. These analyses elucidate the genome structure and gene regulation of fission yeast and provide tools for investigation across the *Schizosaccharomyces* clade.

The fission yeast genus *Schizosaccharomyces* forms a broad and ancient clade within the Ascomycete fungi (Fig. 1A) with a distinct life history from other yeasts (*1*). Fission yeast grow preferentially as haploids, divide by medial fission rather than asymmetric budding, and have evolved a single-celled life-style independently from the budding yeasts (Saccharomycotina). Fission yeasts share important biological processes with metazoans, including chromosome structure and metabolism—relatively large chromosomes, large repetitive centromeres, low-complexity replication origins, heterochromatic histone methylation, chromodomain heterochromatin proteins, small interfering RNA (siRNA)–regulated heterochromatin, and TRF family telomere-binding proteins—G_2/M cell cycle control, cytokinesis, the mitochondrial translation code, the RNA interference (RNAi) pathway, the signalosome pathway, and spliceosome components. These features are absent or highly diverged in budding yeast. In general, core orthologous genes in fission yeast more closely resemble those of metazoans than do those of other Ascomycetes (*2*). Fission yeasts have also evolved innovations in carbon metabolism, including aerobic fermentation of glucose to ethanol (*3*). This convergent evolution with the budding yeast *Saccharomyces cerevisiae* offers insight into the evolution of complex phenotypes.

S. pombe is widely used as a model for basic biological processes in the cell and to study genes implicated in human disease. To better understand its evolution and natural history, we have compared the genomes and transcriptomes of *S. pombe*, *S. japonicus*, *S. octosporus*, and *S. cryophilus*, which constitute all known fission yeasts.

Genome sequence and phylogeny. We sequenced and assembled the genomes of *S. octosporus*, *S. cryophilus*, and *S. japonicus* using clone-based and clone-free whole-genome shotgun (WGS) approaches (table S1). Each genome is ~11.5 Mb in size. *S. octosporus* and *S. cryophilus* are 38% GC; *S. japonicus* is 44%. By comparison, the *S. pombe* genome is 12.5 Mb in size and 36% GC. We assembled the *S. octosporus* and *S. japonicus* scaffolds into three full-length chromosomes of quality similar to that of the finished *S. pombe* genome (Fig. 1B, figs. S1 and S2, and tables S2 and S3) and identified telomeric sequence using WGS data (*4*). Telomere repeats in *S. japonicus* (GTCTTA), *S. octosporus* (GGGTTACTT), and *S. cryophilus* (GGGTTACTT) matched a one-and-a-half repeat–unit sequence at the putative telomerase-RNA locus, similar to the configuration in *S. pombe* (GGTTAC) (*5*). Using these motifs, we extended the *S. japonicus* and *S. octosporus* chromosomes into subtelomeric and telomeric sequence (*4*).

We constructed a phylogeny of the Schizosaccharomycetes within Ascomycota (Fig. 1A and fig. S3) from 440 single-copy core orthologs, placing the monophyletic *Schizosaccharomyces* species as a basal sister group to the clade, including the filamentous fungi (Pezizomycotina) and budding yeast (Saccharomycotina). We found an average amino acid identity of 55% between all 1:1 orthologs when we compared *S. pombe* and *S. japonicus*, similar to that between humans and the cephalochordate amphioxus (table S4). For the most closely related species, *S. cryophilus* and *S. octosporus*, 1:1 orthologs share 85% iden-

[1]Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, 364 Plantation Street, Worcester, MA 01605, USA. [2]Broad Institute of Massachusetts Institute of Technology and Harvard, 320 Charles Street, Cambridge, MA 02141, USA. [3]Broad Institute of Massachusetts Institute of Technology and Harvard, 7 Cambridge Center, Cambridge, MA 02142, USA. [4]Department of Biology, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA. [5]School of Computer Science and Engineering, Hebrew University, Jerusalem 91904, Israel. [6]Department of Microbiology and Molecular Genetics, Faculty of Medicine, Hebrew University, Jerusalem 91120, Israel. [7]Department of Systems Biology, Harvard Medical School, 200 Longwood Avenue, Alpert 536, Boston, MA 02115, USA. [8]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 32 Vassar Street 32-D510, Cambridge, MA 02139, USA. [9]Microbial Genetics Laboratory, Genetic Strains Research Center, National Institute of Genetics, 1111 Yata, Mishima, Shizuoka 411-8540, Japan. [10]Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892, USA. [11]Wellcome Trust Centre for Cell Biology, Institute of Cell Biology, School of Biological Sciences, The University of Edinburgh, 6.34 Swann Building, Mayfield Road, Edinburgh EH9 3JR, UK. [12]Department of Molecular Genetics and Microbiology, Life Science, Room 130, State University of New York, Stony Brook, NY 11794, USA. [13]Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR 97331, USA. [14]Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA. [15]Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA, UK. [16]Centre for Genetics and Genomics, University of Nottingham, Queen's Medical Centre, Nottingham NG7 2UH, UK. [17]Center for Biosciences, Department of Biosciences and Nutrition, Karolinska Institute, 141 38 Huddinge, Sweden. [18]Department of Life Sciences, Södertörns Högskola, 141 89 Huddinge, Sweden. [19]Warwick Medical School, University of Warwick, Gibbet Hill Campus, Coventry CV4 7AL, UK. [20]Broad Institute of Massachusetts Institute of Technology and Harvard, 301 Binney Street, Cambridge, MA 02141, USA. [21]Alexander Silberman Institute of Life Sciences, Hebrew University, Jerusalem 91904, Israel. [22]Stowers Institute for Medical Research, Kansas City, MO 64110, USA. [23]Department of Molecular and Integrative Physiology, University of Kansas Medical School, Kansas City, KS 66160, USA. [24]Howard Hughes Medical Institute.

*Present address: National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Department of Health and Human Services, 45 Center Drive, Bethesda, MD, 20892, USA.

†Present address: Wellcome Trust Centre for Gene Regulation and Expression, College of Life Sciences, University of Dundee, Dundee DD1 5EH, Scotland, UK.

‡Present address: Evolva Biotech A/S, Bülowsvej 25, 1870 Frederiksberg C, Denmark.

§Present address: The University of Arizona, The School of Plant Sciences, 303 Forbes Building, 1140 East South Campus Drive, Tucson, AZ 87421, USA.

‖These authors made equivalent contributions.

¶To whom correspondence should be addressed. E-mail: nick.rhind@umassmed.edu (N.R.); aregev@broad.mit.edu (A.R.); chad@broadinstitute.org (C.N.)

tity on average, similar to humans and dogs. The genetic diversity within *S. pombe* is low. Comparing the *S. pombe* 972 strain to WGS analysis of *S. pombe* NCYC132 and *S. pombe* var. *kambucha*, two phenotypically distinct strains, revealed less than 1% nucleotide difference between the three strains (fig. S4 and table S5).

Eradication of transposons and reorganization of centromere structure. Transposons and other repetitive sequences are thought to be crucial for centromeric function through the maintenance of heterochromatin (*6*). These sequences evolve rapidly, but the evolutionary relation among centromeres, transposons, and heterochromatin is unclear, in part because fungal centromeres have not generally been included in genome assemblies. The *S. japonicus* genome harbors 10 families of gypsy-type retrotransposons (*4*) (fig. S5 and table S6). Sequence divergence of their reverse transcriptases suggests that these transposon families predate the last common ancestor of the Ascomycetes. However, a dramatic loss of transposons occurred after the divergence of *S. japonicus*; *S. pombe* harbors two related retrotransposons, Tf1 and Tf2; *S. cryophilus* has a single related retrotransposon, Tcry1; *S. octosporus* contains no transposons, but contains sequences related to reverse transcriptase and integrase that may represent extinct transposons (fig. S5 and table S6).

The disappearance of transposons in the evolution of fission yeast species after *S. japonicus* correlates with the appearance of the *cbp1* gene family, which suggests a transition in the control of centromere function. In *S. pombe*, Cbp1 proteins bind centromeric repeats and are required for transposon silencing and genome stability (*7*, *8*). Although described as orthologs of CENP-B, a human centromere-binding protein, Cbp1 proteins apparently evolved independently within the *Schizosaccharomyces* lineage from a domesticated Pogo-like DNA transposase (*9*). The appearance of the *cbp1* gene family also correlates with the switch from RNAi-mediated transposon silencing in *S. japonicus* (see below) to a Cbp1-based mechanism in *S. pombe*, which suggests that this shift to Cbp1-based transposon control allowed the eradication of most transposons from the fission yeast genomes, possibly by promoting recombinational deletion between long terminal repeats (LTRs) (*8*). Furthermore, the *cbp1* family is evolving rapidly (fig. S6), which suggests that Cbp1-based transposon silencing is a *Schizosaccharomyces*-specific innovation that arose after the divergence of *S. japonicus*.

The loss of transposons was accompanied by a substantial reorganization of chromosome architecture that conserves centromere function, which suggests that the evolution of novel

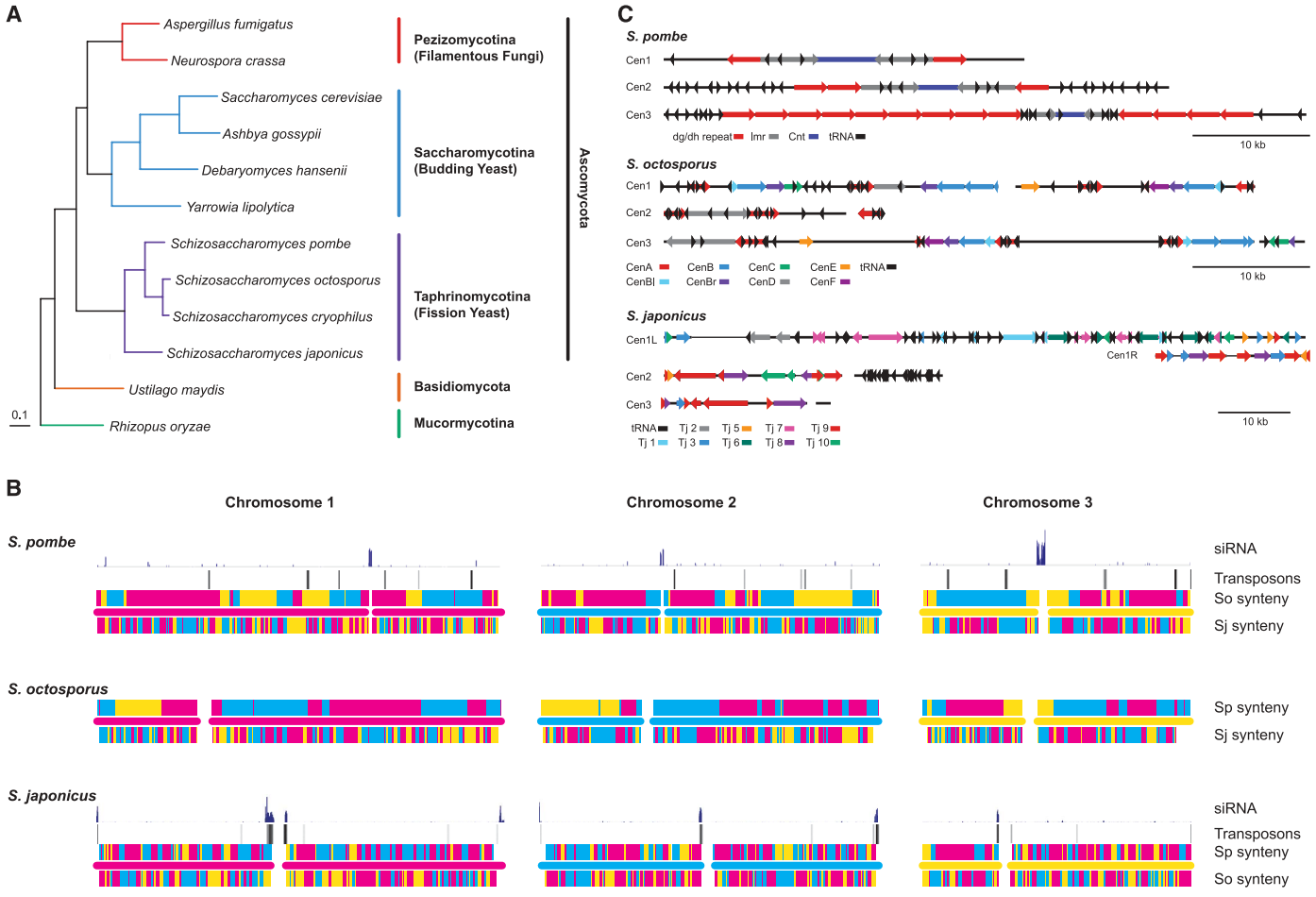


Fig. 1. *Schizosaccharomyces* phylogeny and chromosome structure. (**A**) A maximum-likelihood phylogeny of 12 fungal species from 440 core orthologs (each occurring once in each of the genomes) from fly to yeast. A maximum-parsimony analysis produces the same topology. Both approaches have 100% bootstrap support for all nodes. (**B**) The chromosome structure of *S. pombe*, *S. octosporus*, and *S. japonicus*. The middle bar in each figure represents the chromosome and its centromere: red for chromosome 1, blue for chromosome 2, and yellow for chromosome 3. Depicted above and below each chromosome are the chromosomes in the other two species to which the genes on the chromosome of interest map, using the same color scheme. Depicted above the *S. pombe* and *S. japonicus* chromosomes are the distributions of transposons and mapping of siRNAs. *S. cryophilus* is not included, because its genome has not been assembled into complete chromosomes. (**C**) The centromeric repeat structures of *S. pombe*, *S. octosporus*, and *S. japonicus*. Due to their repetitive nature, they are unlikely to represent the exact genomic structure. The *S. pombe* portion of the figure is adapted from (*11*).

centromere structures compensated for the loss of transposons. In *S. japonicus*, transposons cluster next to telomeres and centromeres, as in metazoans (Fig. 1, B and C). In the other Schizosaccharomycetes, the subtelomeres and pericentromeres are also repetitive, but lack transposons (Fig. 1C). However, like *S. japonicus*, the centromeric and subtelomeric repeats are confined to pericentromeric and subtelomeric regions, respectively, with one exception—a centromeric repeat involved in transcriptional silencing at the *S. pombe* mating-type locus (*10*). We confirmed that the centromeres are heterochromatic by histone H3 lysine 9 (H3K9) methylation mapping (fig. S7) and showed that the *S. japonicus* centromeres are functional by meiotic mapping (table S2).

Although centromeric repeats evolve rapidly, differing even between related strains (*11*), individual repeat sequences tend to be similar within strains (Fig. 1C). No similarity was observed between the centromeric repeats of *S. pombe*, *S. octosporus*, or *S. cryophilus*. However, both *S. pombe* and *S. octosporus* centromeres contain repeated elements, highly similar between chromosomes, that are arrayed in a larger inverted repeat structure around a unique core sequence (Fig. 1C), which suggests that they are homogenized by nonreciprocal recombination. This contrasts with a lack of symmetry in *S. japonicus* and implies that transposition occurs more rapidly than homogenization by recombination. Thus, the suppression of transposition likely led both to the degeneration of transposon sequences and to the evolution of symmetric centromeric repeats.

Despite the divergence of centromere sequence and of gene order on the chromosome arms, karyotype and pericentromeric gene order are conserved between *S. pombe* and *S. octosporus* (fig. S8). Thus, although gene conversion maintains the similarity of centromeric repeats between the different centromeres, crossover recombination between centromeres is suppressed. We observed neither centromeric translocations nor neocentromere events within these lineages, even though centromeres can occur at novel locations in manipulated *S. pombe* strains (*12*). The retention of repetitive elements in the centromeres of *S. pombe*, *S. octosporus*, and *S. cryophilus*, even as they have lost their transposons, implies that centromeric repeats have an important function.

Because siRNAs are involved in both transposon silencing and centromere function (*13*), we investigated these roles in the *Schizosaccharomyces* lineage. In *S. pombe*, the centromeric repeats produce dicer-dependent siRNAs required for maintenance of centromeric structure, function, and transcriptional silencing via Argonaute-dependent heterochromatin formation (*14*). However, transposons are silenced in *S. pombe* by RNAi-independent mechanisms and do not produce abundant siRNAs (Fig. 1B and fig. S9) (*7*). To investigate whether centromere-directed siRNA production is conserved within the transposon-rich centromeres of *S. japonicus*, we sequenced small RNAs from log-phase *S. japonicus* cultures (which have a modal size of 23 nucleotides) (*4*) and found that 94% map to transposons, both telomeric and centromeric (Fig. 1B and fig. S9). The fact that siRNAs map to transposons in *S. japonicus* but not in *S. pombe* suggests that either the fission yeast RNAi pathway targets repetitive sequences instead of mobile elements per se or that the pathway evolved away from an ancestral role in transposon control to a dedicated role in heterochromatin function.

Evolution of mating-type loci. The structure of the mating-type loci and the cis-acting elements that regulate mating-type switching is highly conserved across all four species (fig. S10). The expressed *mat1* locus can contain either the plus (P) or the minus (M) allele and switches between the two by epigenetically programmed gene conversion (*15*–*17*) from two heterochromatically silenced donor cassettes: *mat2-P* and *mat3-M* (figs. S10 and S11). cis-Acting regulatory sequences required for epigenetic imprinting and recombinational switching (*18*–*20*) are conserved (*4*) (fig. S11), as is the epigenetically programmed genomic mark associated with *mat1* (*15*).

In contrast, none of the cis-acting sequences involved in transcriptional repression of the silent cassettes in *S. pombe* are identifiable in the other species, although the donor cassettes are enriched for H3K9me heterochromatin (fig. S7). In *S. japonicus*, the silent *mat* loci are found to directly abut the centromere of chromosome 3, which suggests that they may be silenced by a positional effect. In *S. octosporus* and *S. cryophilus*, the *mat* loci are distant from the centromeres, but each contains a conserved region of transposon remnants, which may be silencing triggers. They also contain inverted repeats, albeit shorter and less similar to each other than the inverted repeats that flank the *mat2/3* locus in *S. pombe* (*21*). Thus, their silencing strategies may share elements from both *S. pombe* and *S. japonicus*. These results suggest that the mechanisms of imprinting and switching have been conserved, but that the strategies for establishing heterochromatin are plastic.

Comparative annotation of transcriptomes. We annotated the three genomes using standard methods and compared them with *S. pombe* (*4*). We then deep-sequenced poly(A)-enriched, strand-specific cDNA (*22*–*24*) (RNA-Seq) and constructed de novo transcript models (fig. S12) for log phase, glucose depletion, early stationary phase, and heat shock from *S. pombe*, *S. octosporus*, and *S. japonicus* and log phase, glucose depletion, and heat shock from *S. cryophilus*.

In *S. pombe*, we reconstructed 4277 out of 5064 previously annotated genes; of the remain-

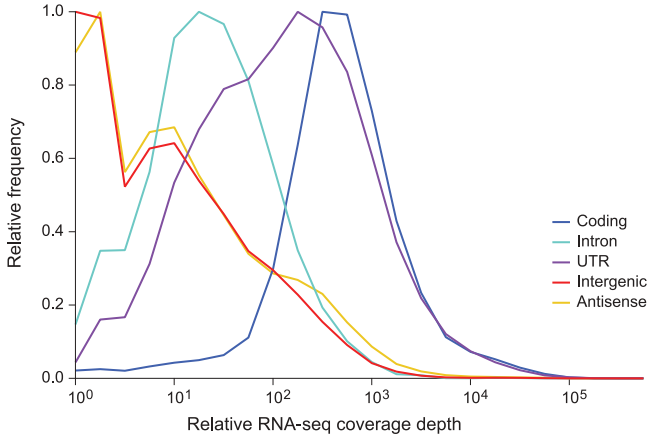


Fig. 2. Polyadenylated transcription is predominantly confined to protein-coding genes. The *S. pombe* genome was divided into four different feature classes: protein-coding sequence, intron sequence, untranslated sequence (5′ and 3′ UTRs), and intergenic sequence (all nucleotides between UTRs of protein-coding genes). The frequency of RNA-Seq reads was calculated over sequential 20–base pair windows across these features; for coding sequence, the frequency of antisense reads was also calculated. Frequency was normalized to the maximum frequency within each feature class to compensate for the different class sizes.

Table 1. Conservation of gene content and structure. See Fig. 5 legend for abbreviations.

Organism and total	Orthologous groups				Introns		
	Same	Gain	Loss	Dup	Same	Gain	Loss
S. pombe	4218	321	83	23	2901	297	27
S. octosporus	4218	133	48	5	2901	25	8
S. cryophilus	4218	283	73	11	2901	75	4
Ancestor of *Soct* and *Scry*	4218	103	44	15	2901	396	0
Ancestor of *Spom*, *Soct*, and *Scry*	4218	339	159	29	2901	415	412
S. japonicus	4218	242	0	18	2901	708	214
Ancestor of *Schizosaccharomyces*		640	745				
Total		2061	1152	101		1916	665

ing 788 genes, 60% were covered over at least 90% of their length. Four hundred of our transcript models change coding exon structure of the gene, 95% of which maintained or improved conserved coding capacity (tables S7 and S8 and fig. S12) (*25*). In addition, we identified 253 untranslated region (UTR) introns. Last, we found 89 new protein-coding genes in *S. pombe*, 53 of which are conserved (table S7 and fig. S13). We found no evidence that intron-rich fission yeast genes engage in metazoan-like alternative splicing (*26*). We found evidence for 433 alternative splicing events in *S. pombe* in the form of intron retention and alternative splice-donor or splice-acceptor usage, but no evidence of exon skipping or alternative exons; we found similar levels of splice variants in the other species (table S9). However, because many of these variants disrupt the coding capacity (figs. S14 and S15) and only a minority of intron retentions (146 out of 393) are conserved between two or more species, we suspect that much of alternative splicing in fission yeast represents nonproductive splicing variants. It is interesting that, in some cases, the nonspliced variant may be the protein-coding isoform (figs. S14, C and D, and S15, and table S10).

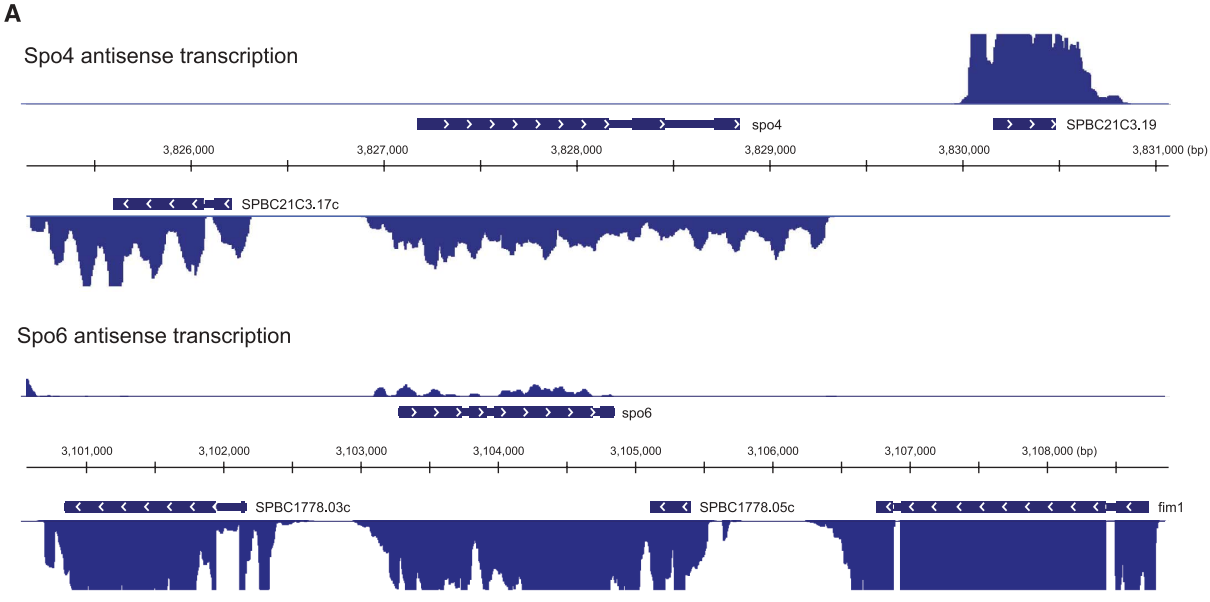
Transcription primarily represents protein-coding transcripts. The majority of stable fission yeast transcripts originate from annotated protein-coding genes. Most of the *S. pombe* genome is transcribed (*22*), with 91% of nucleotides covered by at least one RNA-Seq read. However, most transcription, as measured by steady-state poly(A)-enriched RNA levels, is associated with well-defined transcripts, most of which are protein coding. Specifically, 37% of intergenic nucleotides (between the UTRs of annotated protein-coding transcripts) are not detectibly expressed, and 90% of transcribed intergenic nucleotides account for only 0.16% of the poly(A)-enriched transcript signal. Moreover, the median expression level of exonic sequence (99.1% of which are detectably expressed) is 305 times that of intergenic sequence (Fig. 2 and table S11), with intergenic transcription enriched within origins of DNA replication (fig. S16)—gene-free loci with nucleosome-free regions (*27*–*29*) that may provide permissive loci for ectopic transcriptional initiation (*30*).

Transcription of coding genes is heavily biased to the sense strand. Of the coding genes, 73% have <5% of their RNA-Seq reads on the antisense strand. Genes with >5% antisense reads are enriched for convergent transcripts with intergenic distances of <200 bp ($P < 10^{-8}$, hypergeometric test), but not with those of >200 bp ($P > 0.1$), which suggests that much antisense transcription is due to readthrough of 3′-termination sites (*31*) (fig. S17). Thus, stable transcripts in fission yeast genomes are primarily associated with known transcription units. We discuss notable exceptions below.

Conservation of gene content and structure. Despite the evolutionary breadth of the fission-yeast clade, as measured by amino acid divergence, their gene content and structure are remarkably conserved. Of ~5000 coding genes in fission yeast species, 4218 are 1:1:1:1 orthologs across the clade, with the remainder of the orthologous groups containing genes that have been duplicated or deleted since their last common ancestor (Table 1 and fig. S12). Protein kinases are even more conserved in gene content; 93% (102 out of 110) of *S. pombe* protein kinases are 1:1:1:1 orthologs (*4*). Moreover, of 3601 *S. pombe* introns in 2616 spliced 1:1:1:1 orthologs, 2901 (81%) are identical across the four species (table S13). Overall, the conservation of gene content, gene order, and gene structure within *Schizosaccharomyces* is higher than expected given the level of amino acid divergence. From amino acid divergence, we estimate that the fission yeast clade arose about 220 million years ago (fig. S3). However, the conservation of gene content is significantly higher than that within *Saccharomyces* or *Kluyveromyces*, both of which have much lower amino acid divergence (table S15), which suggests that fission yeast amino acid sequences are evolving anomalously quickly or that genome structures are unusually stable.

The majority of gene changes are due to the gain of species- and clade-specific genes (table S12 and S14) (*4*). We tested whether gene gain is due to rapid divergence of orthologous genes by

A



B

S. pombe GO annotation	p value
Meiosis	1.33×10^{-10}
Meiotic chromosome segregation	1.53×10^{-8}
Meiotic recombination	4.38×10^{-6}
Nuclear chromosome	4.54×10^{-3}
Ascospore formation	5.60×10^{-3}

Fig. 3. Meiotic genes are subject to antisense transcription. (**A**) Examples of antisense transcription of meiotic genes. Above and below the chromosome coordinates are the coding sequence annotations on the top and bottom strand, respectively. Above and below these are the strand-specific RNA-Seq read densities on a 0 to 300 scale; signal above 300 is truncated to make the low-amplitude signal visible. (**B**) Enrichment of Gene Ontology (GO) annotations within the set of protein-coding genes with more antisense than sense transcription. All terms with a *P* value of <0.01 are included, except for high-level terms (i.e., biological process and molecular function).

looking for colinearity in regions with species-specific genes, and we examined these regions for signs of sequence similarity. We found that 94 out of 317 *S. pombe*–specific genes are in the same position relative to neighboring genes as genes specific to other species (table S16). Of these, nine show > 15% identity to a cognate gene in another species, which suggests that they are rapidly diverged orthologs (*4*).

We also found 34 *S. pombe* candidates for horizontal gene transfer from bacteria, including two published examples (*4*, *32*, *33*) (table S17), and similar numbers in the other species. Of these, 16 appear to have occurred before the radiation of the clade, and 9 appear to be specific to *S. pombe*.

Evidence for intergenic and antisense noncoding transcripts. We identified 1097 putative transcript models in *S. pombe* supported by strand-specific RNA-Seq data, but containing no obvious coding capacity and having no correspondence to well-defined noncoding RNAs (ncRNAs) (*22*, *24*, *34*) (fig. S18 and tables S18 and S19). Of these potential ncRNAs, 449 are intergenic and 648 are antisense, overlapping a coding gene on the other strand by at least 30%. Of the ncRNAs, 213 overlap an annotated UTR on the same strand, which suggests that they may be alternative UTRs. Nevertheless, the data support 338 of the intergenic and 546 of the antisense ncRNAs as distinct transcripts (*4*).

Of the 338 distinct intergenic ncRNAs in *S. pombe*, 138 are conserved in location in at least one other species (table S41). Moreover, 26 of the intergenic ncRNAs are conserved in sequence, and of these, 9 are conserved in both location and sequence, which suggests that they represent potentially biologically important noncoding RNAs. The transcripts that are conserved in location but not in sequence may represent functional transcripts that have diverged beyond recognition. Of the antisense transcripts, 328 (51%) are conserved across two or more genomes (table S20), which suggests that they are biologically important (*35*).

Antisense regulation of meiotic transcription. Across fission yeast, the ~250 genes with greater

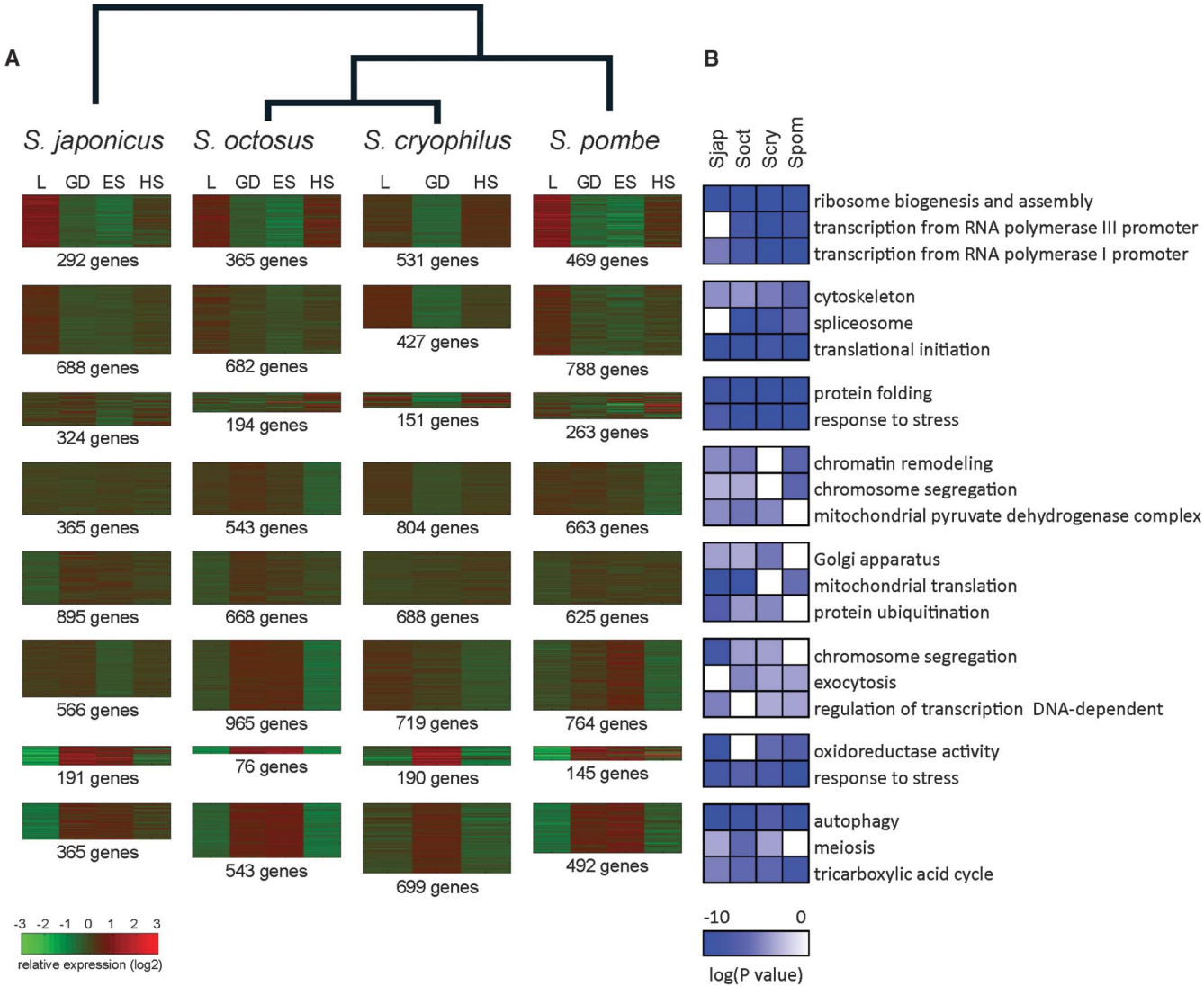


Fig. 4. Expression profiles cluster into similar patterns with conserved biological functions. (**A**) Expression clusters for each species. Gene expression profiles for each species were clustered (*4*). The size of each heat map is proportional to the number of genes in the cluster and the number of genes in each is indicated. Similar cluster sizes and patterns reflect similar expression patterns between the species. The heat shock transcription profile is similar to log-phase growth because the transcriptional response on the 15-min time scale used here is limited to a relatively small number of genes. L, log phase; GD, glucose depletion; ES, early stationary phase; HS, heat shock. (**B**) A selection of enriched GO terms for each cluster. The color intensity is proportional to the negative logarithm of the hypergeometric *P* value enrichment on a continuous scale of 0 to 10. Complete GO term enrichments are shown in table S26.

antisense transcription than sense transcription (table S21) are significantly enriched for meiotic genes ($P = 10^{-10}$ for *S. pombe*, hypergeometric test) (Fig. 3, fig. S19, and tables S22 and S23), consistent with observations in *S. pombe* and *S. cerevisiae* (*24*, *35*). Several antisense-transcribed genes have been proposed to be regulated by intron retention (*36*, *37*); however, these studies did not use strand-specific approaches, which makes it impossible to distinguish unspliced sense transcripts from antisense transcripts. We find no evidence of alternative splicing of any of these genes.

Antisense transcription of meiotic genes does not uniformly decrease as cognate sense transcription increases during meiosis (fig. S20). This observation suggests that antisense transcription does not inhibit sense transcription, in contrast to the anticorrelation observed in *S. cerevisiae* (*30*, *35*). Furthermore, meiotic genes are not enriched among genes with >5% antisense transcription but <100% antisense transcription ($P = 0.47$, hypergeometric test), consistent with a stoichiometric mechanism of regulation in which antisense transcripts directly bind to and inhibit the stability or translation of sense transcripts.

Global conservation of expression programs within fission yeasts. To identify conserved modules of coexpressed genes, we examined expression patterns across the four conditions and between the four fission yeast with phylogenic clustering (Fig. 4). We found that patterns of gene expression between species grown in similar conditions are generally conserved, with dominant patterns associated with growth (log phase and heat shock) and stress (glucose depletion and early stationary phase). Moreover, similar expression clusters are enriched for similar gene annotations across the species.

Fission yeast up-regulate genes involved in mitosis, including those involved in the kinetochore, the spindle pole body, and the anaphase-promoting complex, in response to glucose depletion (table S24). In contrast, several classes of genes involved in growth are down-regulated (*4*). None of these genes are extensively regulated in glucose depletion in *S. cerevisiae* (*38*).

cis-Regulatory mechanisms are associated with novel and expanded functions. Promoter motifs with conserved regulatory function across Ascomycota show new functionality among the *Schizosaccharomyces*. For example, the motif bound by Rtg3 in *S. cerevisiae* is associated with amino acid metabolism genes across the phylum. In fission yeast, however, it is also enriched in genes responsive to various stress responses (Fig. 5A). Of the stress genes that have Rtg3 motifs in *S. pombe*, 36% are found only in the *Schizosaccharomyces* clade, and many are also associated with the Atf1 motif, a conserved regulator of the stress response (Fig. 5B). Rtg3 does not have a detectable ortholog in the *Schizosaccharomyces*

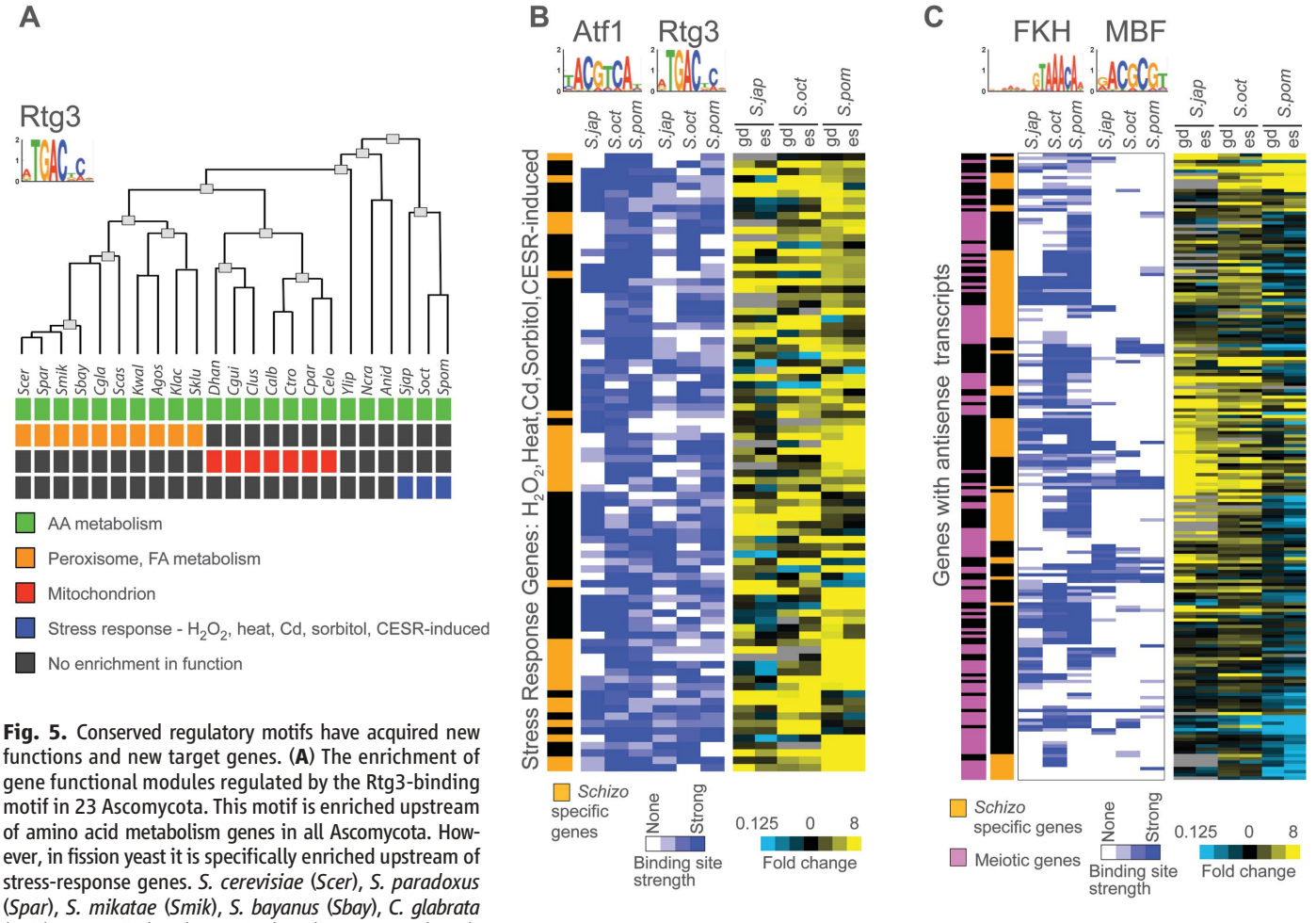


Fig. 5. Conserved regulatory motifs have acquired new functions and new target genes. (**A**) The enrichment of gene functional modules regulated by the Rtg3-binding motif in 23 Ascomycota. This motif is enriched upstream of amino acid metabolism genes in all Ascomycota. However, in fission yeast it is specifically enriched upstream of stress-response genes. *S. cerevisiae* (*Scer*), *S. paradoxus* (*Spar*), *S. mikatae* (*Smik*), *S. bayanus* (*Sbay*), *C. glabrata* (*Cgla*), *S. castellii* (*Scas*), *K. waltii* (*Kwal*), *A. gossypii* (*Agos*), *K. lactis* (*Klac*), *S. kluyveri* (*Sklu*), *D. hansenii* (*Dhan*), *C. guilliermondii* (*Cgui*), *C. lusitaniae* (*Clus*), *C. albicans* (*Calb*), *C. tropicalis* (*Ctro*), *C. parapsilosis* (*Cpar*), *C. elongosporus* (*Celo*), *Y. lipolytica* (*Ylip*), *N. crassa* (*Ncra*), *A. nidulans* (*Anid*), *S. japonicus* (*Sjap*), *S. octosporus* (*Soct*), and *S. pombe* (*Spom*). (**B**) Enrichment of Rtg3- and Atf1-binding sites in the promoters of stress-response genes. Each row represents a gene. The strength of the strongest regulatory site upstream of the gene is indicated in the blue heat map. The expression of the gene in glucose depletion (gd) and early-stationary phase(es) relative to log phase is indicated in the blue-yellow heat map. Genes specific to the fission yeast clade are indicated in orange. (**C**) Enrichment of FKH- and MBF-binding sites in front of antisense-transcribed genes. As in (B), but each row represents a gene with greater antisense than sense transcription. Genes associated with meiosis (*44*) are indicated in magenta. CESR, core environmental stress response.

clade (*39*), but the motif recognized by Rtg3 in *S. cerevisiae* is clearly identifiable in fission yeast, which suggests that these regulatory motifs are more conserved than their binding proteins. We also found a similar acquisition of *Schizosaccharomyces*-specific genes by the FKH- and MBF-associated motifs, which regulate meiotic transcription in *S. pombe* (*4*, *40*, *41*). In particular, these two motifs were enriched in genes with antisense transcripts (Fig. 5C). Most of the FKH (a motif bound by Mei4 in *S. pombe*) target genes with antisense transcripts (80%, 47 genes) are meiotic genes, the majority of which are specific to the *Schizosaccharomyces* clade (Fig. 5C).

Gene content reflects glucose-dependent lifestyle. Fission yeast and budding yeast of the *Saccharomyces* clade independently evolved the ability to produce ethanol by aerobic fermentation (*3*, *42*). In contrast to the convergent evolution of ethanol production, the utilization of ethanol has not converged; although budding yeast can efficiently catabolize ethanol, fission yeast cannot use ethanol as a primary carbon source. The evolution of aerobic fermentation in budding yeast involved changes in gene content, most notably following a whole-genome duplication (WGD) event and in regulatory mechanisms of glucose repression (*3*, *43*).

Like budding yeast, fission yeast have duplicate copies of the pyruvate decarboxylase (*pdc*) gene, needed to funnel pyruvate to fermentation. They also have orthologs of several activators and repressors of respiratory genes, including Hap2/3/4/5 complex members, the Adr1, Tup, and Mig transcriptional regulators, and the Snf1–Sip1 and 2 kinase (*3*). However, there are substantial distinctions in gene content between fission yeast and the post-WGD budding yeast (fig. S21). We identified loss of the glyoxylate cycle, loss of the glycogen biosynthesis, fewer glycolytic paralogs, loss of the gluconeogenic enzyme phosphoenolpyruvate carboxykinase, lack of expanded *adh* genes, and lack of transcriptional regulators of glucose repression as differences that illuminate the distinct metabolic capacities of fission yeast (*4*). All of these adaptations are consistent with the inability of fission yeast to consume ethanol as a sole carbon source. The loss of conserved enzymes highlights how fission yeast came to depend solely on glucose.

In both fission yeast and budding yeast, as glucose is depleted, the expression of respiratory genes [oxidative phosphorylation enzymes or tricarboxylic acid (TCA) cycle] is induced. However, unlike *S. cerevisiae* (*38*), in fission yeast the expression of the genes encoding the pyruvate dehydrogenase complex and *adh1* is reduced, which prevents the efficient use of pyruvate for respiration. Instead, the expression of the *ald* genes is induced, which may provide an alternative mechanism for generating acetyl–coenzyme A in fission yeast.

Thus, the lack of efficient ethanol catabolism by fission yeast demonstrates that aerobic fermentation did not evolve to create a consumable by-product. Instead, ethanol is a waste product, possibly produced because it is toxic to competing microorganisms. It is interesting that aerobic fermentation appears to have evolved as early as 200 million years ago in fission yeast (fig. S3), long before the WGD and subsequent evolution of aerobic fermentation in budding yeast.

Conclusions. Our comparative analysis of genome structure and expression in the fission yeast, especially the analysis of centromere structure and evolution, demonstrates how chromosomal features can be rearranged while retaining function and maintaining stable positions across taxa. We also provide insight into centromeric biology and elucidate conserved antisense transcription that may play a systematic role in meiotic gene regulation. Last, this study informs our understanding of the major evolutionary innovation of aerobic alcohol fermentation in microbial metabolism that arose in parallel in the fission yeast and budding yeast lineages. As these results demonstrate, comparative analyses improve the power of fission yeast as a model for eukaryotic biology.

References and Notes

1. S. L. Forsburg, *Trends Genet.* **15**, 340 (1999).
2. V. Wood *et al.*, *Nature* **415**, 871 (2002).
3. C. L. Flores, C. Rodríguez, T. Petit, C. Gancedo, *FEMS Microbiol. Rev.* **24**, 507 (2000).
4. Materials and methods are available as supporting material on *Science* Online.
5. J. Leonardi, J. A. Box, J. T. Bunch, P. Baumann, *Nat. Struct. Mol. Biol.* **15**, 26 (2008).
6. L. H. Wong, K. H. Choo, *Trends Genet.* **20**, 611 (2004).
7. H. P. Cam, K. Noma, H. Ebina, H. L. Levin, S. I. Grewal, *Nature* **451**, 431 (2008).
8. M. Zaratiegui *et al.*, *Nature* **469**, 112 (2011).
9. C. Casola, D. Hucks, C. Feschotte, *Mol. Biol. Evol.* **25**, 29 (2008).
10. S. I. Grewal, A. J. Klar, *Genetics* **146**, 1221 (1997).
11. N. C. Steiner, K. M. Hahnenberger, L. Clarke, *Mol. Cell. Biol.* **13**, 4578 (1993).
12. K. Ishii *et al.*, *Science* **321**, 1088 (2008).
13. S. I. Grewal, *Curr. Opin. Genet. Dev.* **20**, 134 (2010).
14. T. A. Volpe *et al.*, *Science* **297**, 1833 (2002).
15. D. H. Beach, *Nature* **305**, 682 (1983).
16. R. Egel, *Curr. Genet.* **8**, 205 (1984).
17. S. Vengrova, J. Z. Dalgaard, *Genes Dev.* **18**, 794 (2004).
18. S. Sayrac, S. Vengrova, E. L. Godfrey, J. Z. Dalgaard, *PLoS Genet.* **7**, e1001328 (2011).
19. M. Kelly, J. Burke, M. Smith, A. Klar, D. Beach, *EMBO J.* **7**, 1537 (1988).
20. B. Arcangioli, A. J. Klar, *EMBO J.* **10**, 3025 (1991).
21. G. Singh, A. J. Klar, *Genetics* **162**, 591 (2002).
22. B. T. Wilhelm *et al.*, *Nature* **453**, 1239 (2008).
23. J. Z. Levin *et al.*, *Nat. Methods* **7**, 709 (2010).
24. T. Ni *et al.*, *PLoS ONE* **5**, e15271 (2010).
25. M. F. Lin, I. Jungreis, M. Kellis, *Nature Precedings*, published online 18 August 2010 (http://hdl.handle.net/10101/npre.2010.4784.1).
26. N. F. Käufer, J. Potashkin, *Nucleic Acids Res.* **28**, 3003 (2000).
27. M. Gómez, F. Antequera, *EMBO J.* **18**, 5683 (1999).
28. M. L. Eaton, K. Galani, S. Kang, S. P. Bell, D. M. MacAlpine, *Genes Dev.* **24**, 748 (2010).
29. A. B. Lantermann *et al.*, *Nat. Struct. Mol. Biol.* **17**, 251 (2010).
30. Z. Xu *et al.*, *Nature* **457**, 1033 (2009).
31. M. Zofall *et al.*, *Nature* **461**, 419 (2009).
32. T. Matsuzawa *et al.*, *Appl. Microbiol. Biotechnol.* **87**, 715 (2010).
33. T. Uo, T. Yoshimura, N. Tanaka, K. Takegawa, N. Esaki, *J. Bacteriol.* **183**, 2226 (2001).
34. N. Dutrow *et al.*, *Nat. Genet.* **40**, 977 (2008).
35. M. Yassour *et al.*, *Genome Biol.* **11**, R87 (2010).
36. A. Moldón *et al.*, *Nature* **455**, 997 (2008).
37. N. Averbeck, S. Sunder, N. Sample, J. A. Wise, J. Leatherwood, *Mol. Cell* **18**, 491 (2005).
38. J. L. DeRisi, V. R. Iyer, P. O. Brown, *Science* **278**, 680 (1997).
39. I. Wapinski, A. Pfeffer, N. Friedman, A. Regev, *Nature* **449**, 54 (2007).
40. N. F. Lowndes, C. J. McInerny, A. L. Johnson, P. A. Fantes, L. H. Johnston, *Nature* **355**, 449 (1992).
41. H. Abe, C. Shimoda, *Genetics* **154**, 1497 (2000).
42. J. Piskur, E. Rozpedowska, S. Polakova, A. Merico, C. Compagno, *Trends Genet.* **22**, 183 (2006).
43. M. Kellis, N. Patterson, M. Endrizzi, B. Birren, E. S. Lander, *Nature* **423**, 241 (2003).
44. J. Mata, R. Lyne, G. Burns, J. Bähler, *Nat. Genet.* **32**, 143 (2002).

Acknowledgments: Assemblies and annotations are available at GenBank (*S. octosporus*, ABHY04000000; *S. cryophilus*, ACQJ02000000; *S. japonicus*, AATM02000000), the Broad Institute's *Schizosaccharomyces* Web site (www.broadinstitute.org/annotation/genome/schizosaccharomyces_group), which provides search and visualization tools and pomBase (www.pombase.org). The RNA-Seq and single-nucleotide polymorphism (SNP) data are at the National Center for Biological Information (NCBI), NIH, Sequence Read Archive (see table S42). The *S. japonicus* siRNA data sets are at NCBI GEO as GSE26902 and GSE27837. This work was supported by the National Human Genome Research Institute (NHGRI), NIH. C.N. and M.F.L. were supported by NHGRI; M.Y. was supported by a Clore Fellowship; I.W. is the Howard Hughes Medical Institute (HHMI) fellow of the Damon Runyon Cancer Research Foundation; S.R. was supported by NSF; R.M. was supported by NIH; K.H. was supported by Danish Research Council; C.A.N and C.A.M were supported by the Biotechnology and Biological Science Research Council; P.B. was supported by the Stowers Institute and HHMI; Y.G. and H.L. were supported by the National Institute of Child Health and Human Development, NIH; M.K. was supported by the NIH, an NSF CAREER award, and the Sloan Foundation; A.R. was supported by the Human Frontier Science Program, a Career Award at the Scientific Interface from the Burroughs Wellcome Fund, the Sloan Foundation, an NIH Director's PIONEER award and HHMI. We thank the Broad Institute Sequencing Platform; A. Fujiyama and A. Toyoda for generating DNA sequence; M. Lara and N. Stange-Thomann for developing molecular biology protocols; J. Robinson, M. Garber, and P. Muller for technical advice and support; A. Klar for providing *S. pombe* var. *kambucha* (SPK1820); L. Gaffney for assistance with the figures; K. Mar and J. Mwangi for administrative support; and C. Cuomo for comments on the manuscript.

Supporting Online Material

www.sciencemag.org/cgi/content/full/science.1203357/DC1
Materials and Methods
Figs. S1 to S25
Tables S1 to S42
References

26 January 2011; accepted 7 April 2011
Published online 21 April 2011;
10.1126/science.1203357